Mitigation of Human Supervisory Control Wait Times through Automation Strategies

by

Paul Jeffrey Mitchell

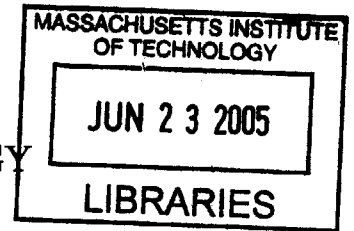B.S., Queen's University, Kingston, Ontario, Canada, 2002

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2005   [June 2005]

Author ....
Department of Aeronautics and Astronautics
May 20, 2005

Certified by.................................
Mary ((Missy) L. Cummings
Boeing Assistant Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by.............
Jaime Peraire
Professor of Aeronautics and Astronautics
Chair, Committee on Graduate Students

AERO

Mitigation of Human Supervisory Control Wait Times through Automation Strategies

by

Paul Jeffrey Mitchell

Abstract

The application of network centric operations principles to human supervisory control (HSC) domains means that humans are increasingly being asked to manage multiple simultaneous HSC processes. However, increases in the number of available information sources, volume of information and operational tempo, all which place higher cognitive demands on operators, could become constraints limiting the success of network centric processes. In time-pressured scenarios typical of networked command and control scenarios, efficiently allocating attention between a set of dynamic tasks is crucial for mission success. Inefficient attention allocation leads to system wait times, which could eventually lead to critical events such as missed times on targets and degraded overall mission success. One potential solution to mitigating wait times is the introduction of automated decision support in order to relieve operator workload. However, it is not obvious what automated decision support is appropriate, as higher levels of automation may result in a situation awareness decrement and other problems typically associated with excessive automation such as automation bias.

To assess the impact of increasing levels of automation on human and system performance in a time-critical HSC multiple task management context, an experiment was run in which an operator simultaneously managed four highly autonomous unmanned aerial vehicles executing an air tasking order, with the overall goal of destroying a pre-determined set of targets within a limited time period. A 4x2(3) repeated measures design was utilized in which the level of decision support provided to operators was a between-subjects factor and level of re-planning, which represents operational tempo, a within-subjects factor. The automated decision support, which took the basic form of a timeline display to aid with scheduling, had four increasing levels: manual, passive, collaborative, and management-by-exception. Level of re-planning refers to how much operators were required to adjust the initial plan in flight, based on unexpected occurrences such as changing deadlines or target sets, and included low and high levels.

The passive level of decision support, which provided assistance to the user through color coding and re-organization of scheduling information, was the best overall per-

former as it had no major drawbacks. Management-by-exception had the best performance across several metrics but had a greater number of catastrophic events occur where a UAV erroneously destroyed a friendly target. The manual level performed better than expected, but had a similarly high critical event rate. Contrary to expectations, the collaborative level of decision support, which provided predictions for possible periods of task overload as well as possible courses of action to relieve the high workload, had the worst performance. This is attributable to an unintended consequence of the automation where the graphical visualization of the computer's predictions caused users to try to globally optimize the schedules for all UAVs instead of locally optimizing schedules in the immediate future, resulting in them being overwhelmed. Total system wait time across both experimental factors was dominated by wait time caused by lack of situation awareness, which is difficult to eliminate, implying that there will be a clear upper limit on the number of vehicles that any one person can supervise because of the need to stay cognitively aware of unfolding events.

Thesis Supervisor: Mary (Missy) L. Cummings
Title: Boeing Assistant Professor of Aeronautics and Astronautics

Acknowledgments

Thank you to my research advisor, Missy Cummings, for your constant guidance and support. You always believed in me, pushed me to the best of my abilities and allowed me greater than expected freedom in choosing the path I took with my thesis. With a busy schedule and an always expanding set of graduate students to advise, you have always been responsive to my needs and provided many great ideas along the way.

Second, many thanks to my parents for always encouraging me to pursue what I love and to always aim for the very best in whatever I do. Their unwavering support is what convinced me to come to MIT without a guarantee of financial assistance, and has been a solid foundation to lean on while here.

Many thanks to all the Humans and Automation Lab graduate students I've worked and interacted with in the past couple of years: Sylvain Bruni, Chris Tsonis, Angela Ho, Jinho Jang, Cristin Smith, and Jessica Marquez. I thank them all for helpful feedback on my interface and experimental designs, as well as for creating a welcoming and collaborative atmosphere in the lab in which it was a pleasure to work.

I would like to give special thanks to David Moore-Pitman, without whose tireless programming efforts this thesis would never have otherwise been possible. Thanks to Stephane Mercier for his help in refining the performance score detailed in Appendix D. I am also indebted to Tom Sheridan for his always insightful comments and encouragement at every stage of this process.

Thanks to everyone else who made my time at MIT fun and exciting, in particular, Kelly Klima, Bill Nadir, Leeland Ekstrom, and Mike Rinehart. Thank you for the fond memories; I leave MIT not just with a Masters degree, but with lifelong friends as well.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

A	Active (Level of Automation)
ATO	Air Tasking Order
AWC	Air Warfare Coordinator
BDA	Battle Damage Assessment
BDI	Battle Damage Imagery
C^2	Command and Control
CE	Command Expression
CSA	Context Switching and Acquisition
DoD	Department of Defense
DARPA	Defense Advanced Research Projects Agency
ETA	Estimated Time of Arrival
FO	Fan Out
HSC	Human Supervisory Control
HSCMTM	Human Supervisory Control Multiple Task Managment
IT	Interaction Time
LOA	Level of Automation
LORP	Level of Re-planning
M	Manual (Level of Automation)
MAUVE	Multi-Aerial Unmanned Vehicle Experiment
MICA	Mixed-Initiative Control of Automata
MMWS	Multi-Modal Workstation
NASA	National Aeronautics and Space Adminstration

NCW	Network Centric Warfare
NCO	Network Centric Operations
NT	Neglect Time
P	Passive (Level of Automation)
PPS	Planning and Problem Solving
RoE	Rules of Engagement
ROTC	Reserve Officer Training Corps
RTB	Return to Base
SA	Situation Awareness
SA	Super Active (Level of Automation)
SEAD	Suppression of Enemy Air Defenses
SPAWAR	Space and Naval Warfare
TOT	Time on Target
TLX	Task Load Index
UAV	Unmanned Aerial Vehicle
UCAV	Unmanned Combat Aerial Vehicle
UGV	Unmanned Ground Vehicle
USAF	United States Air Force
UUV	Unmanned Underwater Vehicle
VMS	Vehicle Monitoring and Selection
WT	Wait Time
WTT	Total System Wait Time
WTI	Interaction Wait Time
WTQ	Wait Time in the Queue
WTSA	Situation Awareness Wait Time

Chapter 1

Introduction

1.1 Motivation

As modern technology continues to evolve, the role of humans in many systems is shifting from manually controlling a system to that of supervising the system, otherwise known as human supervisory control (HSC). HSC is the process by which a human operator intermittently interacts with a computer, receiving feedback from and providing commands to a controlled process or task environment which is connected to that computer. Figure 1-1, adapted from Sheridan [38], illustrates this concept. Human supervisory control is comprised of five generic functions, usually accomplished in the following cyclical order: planning a computer-based task, teaching the computer what was planned through various inputs, monitoring the computer's actions for errors and/or failures, intervening when the plan has been completed or the computer requires assistance, and then the human learns from the experience [38].
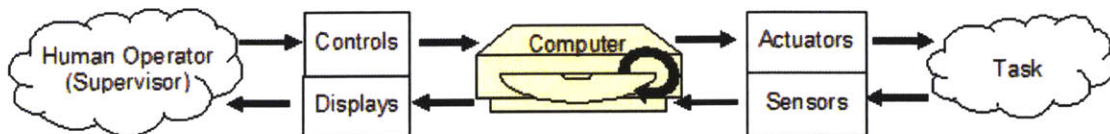
Figure 1-1: Human Supervisory Control [38]

As HSC tasks are primarily cognitive in nature and generally do not require constant attention and/or control, it is possible for humans to effectively supervise multiple simultaneous HSC processes. For example, a single air traffic controller can handle multiple aircraft because the onboard pilots handle the flying task, while the controller is primarily concerned with navigation and deconfliction tasks that do not require constant attention. In order to maximize human performance for the purposes of cost reduction, efficiency, and safety, there is considerable interest in increasing the number and type of HSC processes a single human can handle. Therefore, it is ever more common for humans to be engaged in multiple HSC task management.

1.1.1 Network Centric Operations

Network Centric Warfare (NCW) is a concept of operations envisioned to increase combat power by effectively linking or networking knowledgeable entities in a battlespace. Greater combat power is generated through the creation of shared situational awareness, increased speed of command, self-synchronization, and higher operational tempo, lethality and survivability [1]. NCW's basic tenets (Figure 1-2) are as follows [11]:

1. A robustly networked force improves information sharing.

2. Information sharing and collaboration enhance the quality of information and shared situational awareness.

3. Shared situational awareness enables self-synchronization.

4. These, in turn, dramatically increase mission effectiveness.

A force with these capabilities is therefore able to increase combat power by leveraging information superiority. This is a substantial change from the past when traditional methods for boosting combat power were driven by numerical superiority.

NCW is a broad concept that is applicable to many different HSC domains other than the military, such as commercial aviation, business ventures, and emergency
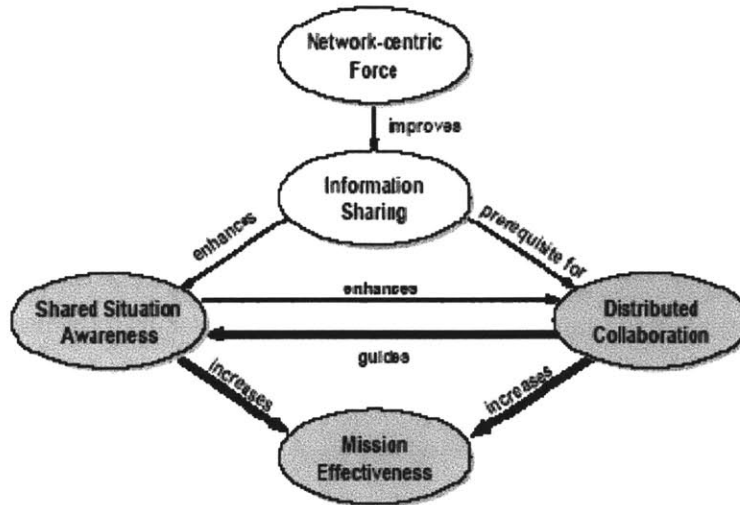
Figure 1-2: Tenets of NCW [23]

response agencies who are also attempting to leverage intelligent information networks to produce superior performance in time critical settings. However, the primary advantage of operations based upon the tenets of NCW (network centric operations or NCO), that of rapid access to information across the network, will likely be a major bottleneck and possible point of failure for those humans who must synthesize voluminous data from the network and execute decisions in real-time, often with high-risk consequences under significant uncertainty. Network-centric operations will bring increases in the number of available information sources, volume of information and operational tempo, all which place higher cognitive demands on operators.

In time-pressured scenarios like those typical of command and control, efficiently allocating attention between sets of dynamic tasks becomes critical to both human and system performance. Inefficient attention allocation could lead to system wait times, which might eventually lead to degraded overall mission success. For example, consider the role of an Air Warfare Coordinator (AWC) on a naval ship engaged in an air defense. The AWC's responsibilities include identifying unknown air tracks, monitoring previously identified tracks, issuing warnings to enemy or unknown aircraft reaching minimum distances from the ship, and providing launch orders for defensive counter measures, if required. This is a HSC system, as the operator is exerting indi-

21

rect control over friendly forces, commercial aircraft and maybe even enemy aircraft in the immediate area of the ship. It is likely that this operator will have many simult... neous tasks to supervise, and if saturated, system wait times could be incurred due to the high workload of the human operator. Inefficient attention allocation could lead to an increase in wait times such that, for example, friendly aircraft receive orders too late to prevent enemy aircraft from attacking, resulting in a damaged or sunk ship. It is also possible that if the human in the decision loop is saturated, disengagement of aircraft or weapons systems prosecuting an inappropriate target such as a commercial aircraft could also be missed.

1.1.2 Unmanned Aerial Vehicles

An unmanned aerial vehicle (UAV) is a powered, aerial vehicle that does not carry a human operator, can fly autonomously or be piloted remotely, is expendable or recoverable, and can carry a lethal or non-lethal payload [10]. They are remotely controlled or autonomous aircraft used primarily in military applications for surveillance or strike missions, and can range in size from handheld miniature vehicles such as the Desert Fox (Figure 1-3) to more long range vehicles such as the Predator (Figure 1-4). The range of commercial applications for UAVs is growing very quickly, as they have found use in border patrol, law enforcement, fire fighting, agriculture, and weather monitoring applications, amongst others. The Aerosonde (Figure 1-5) is an example of a UAV with established uses in weather and atmospheric monitoring.

Unmanned aerial vehicles have been increasingly utilized by militarized forces around the world in recent years to take advantages of their reduced radar signature, increased endurance and decrease in both capital costs and potential loss of human life during operations in hostile territory. The growing importance of UAVs is reflected in the amount of funding allocated to them in the US Department of Defense's (DoD) budget: approximately $1 billion USD in 2004, up from $760 million USD in 2002 and $360 million USD in 2003 [30].

To take full advantage of UAVs in network centric operations, it is desired to have larger "swarm" of vehicles operating in concert. An integral part of this vision involves

Figure 1-3: The Desert Hawk Miniature Unmanned Aerial Vehicle

dramatically changing the number of operators it takes to effectively supervise these vehicles. Multiple operators are currently required to supervise and control the largest and most complex unmanned combat aerial vehicles (UCAVs) such as the Predator, which has a crew of three directly controlling it and a support staff many times this number. In the future, the goal is to invert this ratio so that a single operator supervises the operations of multiple UAVs, and thus this becomes a multiple HSC attention allocation problem. To do so will not only require significant advances in onboard vehicle control and autonomy, but also in the interfaces used for human-computer interaction. Cummings and Guerlain [9] found that Navy personnel were able to effectively manage up to 12 simulated re-targetable Tomahawk missiles, which are highly autonomous and do not require any intervention for in-flight stability control and landing. However, the number of UAVs that can be effectively supervised by operators is likely much less, as they can perform a far wider variety of tasks and usually require some level of flight and navigational control from the human. Ruff et al. [35] studied operators controlling up to 4 autonomous UAVs, and Wickens et al.

23

Figure 1-4: The Predator MQ-1 Unmanned Combat Aerial Vehicle

[44] designed a similar simulator where operators supervised 2 vehicles. Both found that operators could adequately manage these numbers of UAVs so long as they had a basic level of decision support. Significant performance decrements were seen under more manual forms of decision support, which indicates that a major limiting factor on how many vehicles can be effectively supervised is operator workload. As operator workload increases to saturation, the amount of time vehicles have to wait to receive attention when they need it, either for goal state changes or emergency operations, will steadily increase, potentially degrading mission effectiveness.

1.1.3 Automated Decision Support

One potential solution to mitigating system wait times in human supervisory control of UAVs is the introduction of automated decision support in order to relieve operator workload. For example, an accident with a Predator occurred in Bosnia in 1999 simply because the operators experienced an unusual flight condition, aircraft icing, and the added workload associated with this unusual state dramatically increased the system
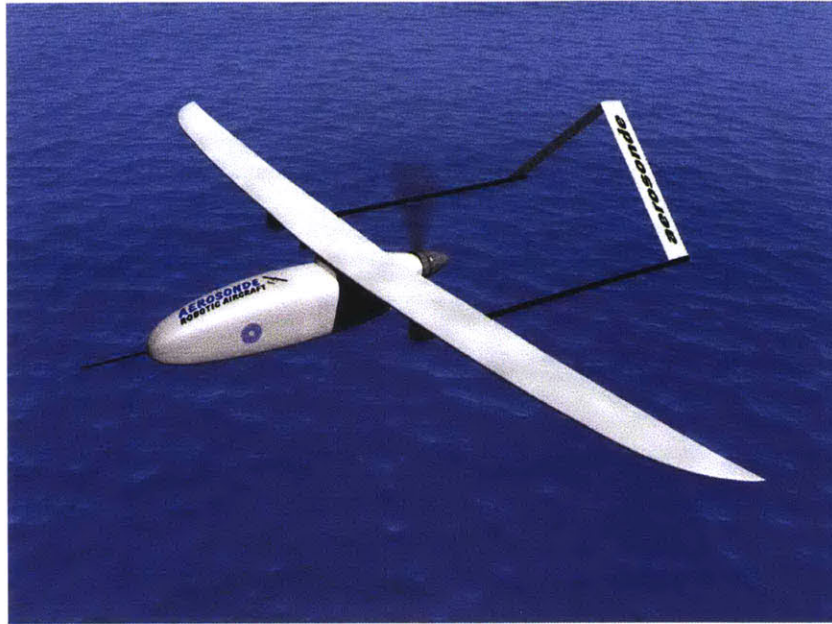
Figure 1-5: The Aerosonde Unmanned Aerial Vehicle

wait time beyond critical levels. Introduction of automated decision support through automatic flight control would have decreased wait times sufficiently such that the accident could have been avoided. However, it is not obvious what automated decision support is appropriate in general for multiple UAV supervision, as higher levels of automation (LOA) may result in a situation awareness decrement and automation bias. This problem has been studied to some degree by Ruff et al. [35] and Wickens et al. [44], but they focused solely on human performance metrics, and did not consider the overall system in their evaluations. In the future vision of allowing a single operator to control multiple unmanned vehicles, it is not well understood how operators will manage multiple vehicles, what kind of decision support will aid or hinder the operator, and how human cognitive limitations will impact overall system effectiveness.

1.2 Problem Statement

The primary questions for this research effort are:

- How do how unmanned vehicle operators cope with managing multiple simultaneous human supervisory control processes, particularly under high workload conditions?

- What amount and types of decision support can best aid operators in these situations?

- What effects do human performance limitations have on system performance?

1.3 Research Objectives

The research objectives of this thesis, in the context of multi-vehicle, time-critical human supervisory control are:

- to develop and validate a model of system wait times,

- to present conclusions on how system wait times and human performance influence each other, and consequently what the implications of this are on the ability of humans to effectively supervise multiple autonomous vehicles,

- to determine what levels of automation are best for supporting operators in terms of both human and system performance characteristics, and

- to study the general cognitive strategies employed by overloaded operators.

1.4 Thesis Organization

This thesis is organized into the following seven chapters:

- Chapter 1, *Introduction*, introduced and described the motivation and objectives for this study.

- Chapter 2, *Background*, provides an introduction to past human factors work in multiple UAV control and supervision, and frames the research objectives outlined above in the context of this prior work. A wait time model is introduced as a metric of system performance and the effect of levels of automation on this wait time model is discussed.

- Chapter 3, *Simulation and Interface Design*, presents the details of an interface and simulation developed to study human and system performance issues involved with the human supervisory control of multiple vehicles in time-critical applications.

- Chapter 4, *Hypotheses and Methods*, discusses predicted system performance trends with increasing levels of automation, and presents the details of an experiment conducted with the program described in the *Simulation and Interface Design* chapter.

- Chapter 5, *Results*, presents the statistical results of the experiment described in the *Methods* chapter.

- Chapter 6, *Discussion*, compares the results with the hypotheses outlined in the Hypotheses and Methods chapter and discusses how they answer the primary research questions of this study.

- Chapter 7, *Conclusion*, summarizes the key findings of this study and presents concluding remarks. Recommendations for future work are also suggested.

Chapter 2

Background

This chapter provides a summary of relevant prior human factors work in multiple vehicle control and supervision from both the human and system perspectives. It then shows how the goals of this research attempt to address the gap between these two approaches to system design.

2.1 Overview

The application of network centric warfare principles to human supervisory control domains means that humans are increasingly being asked to manage multiple simultaneous HSC processes. At the same time, NCO brings greater amounts and variety of information to the operator, placing potentially overwhelming cognitive demands on them. One information overload strategy is to introduce automated systems of varying degrees to relieve operator workload so that overall mission performance is improved. Previous work has been conducted to study human performance and UAV control with varying levels of automation, but these studies did not fully explore the impacts on overall system performance [34, 45]. At the same time, the problem of workload and control of multiple vehicles has been studied from the system perspective without regard for cognitive human-system interaction issues [27]. This chapter discusses these separate approaches and then relates them to each other to identify the gap that this research attempts to address.

2.2 Human Performance

Gawron [16] and Mouloua et al. [25] provide informative summaries of the many human factors issues associated with UAVs. Significant effort has been expended on developing better visualization techniques and interfaces to improve UAV operators' performance and situation awareness (for example, Draper and Ruff [12], Wickens et al. [44]), but very few detailed efforts have investigated how to properly allocate tasks between UAVs and operators, and in particular between multiple UAVs and operators. Current operations of large scale UAVs require many operators per vehicle because of the need for humans to perform low level tasks such as flight control and image processing. However, in the future it is likely the level of autonomy built into such vehicles will increase to the point that the human's role becomes supervisory in nature. As it is desired to eventually invert the ratio of many operators controlling one vehicle to one operator supervising many vehicles, there is a need for greater vehicle autonomy and higher levels of decision support for operators. However, it is not clear what degree of autonomy and/or decision support is optimal to support this goal.

2.2.1 Appropriate Levels of Automation

Automation allocation for decision support can range from fully automatic where the operator is completely left out of the decision process to minimal levels where the automation offers basic data filtering or recommendations for the human to consider. Table 2.1, originally proposed by Sheridan and Verplank [39], outlines a scale commonly used to characterize the allocation of function between man and machine. Human interaction with automation represents a range of intermediate levels from 1-6 on this scale.

For routine operations, higher levels of automation in general result in lower workload, while the opposite is true for low levels of automation [20]. The relationship between workload and performance is illustrated in Figure 2-1, which is adapted to the Yerkes-Dodson Law [46]. This illustration shows that optimal human perfor-

Automation Level	Automation Description
1	The computer offers no assistance: human must take all decision and actions.
2	The computer offers a complete set of decision/action alternatives, or
3	narrows the selection down to a few, or
4	suggests one alternative, and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs humans, and
8	informs the human only if asked, or
9	informs the human only if it, the computer, decides to.
10	The computer decides everything and acts autonomously, ignoring the human.

Table 2.1: Levels of Automation [39]

mance occurs at moderate levels of workload. As a consequence, there is an optimal level of workload, and thus an optimal level of automation whereby performance is maximized for a particular task. Therefore, performance degradation can occur if a LOA for a task is selected that is too high or too low. If the LOA is too high, HSC problems can include: 1) manual or mental skill degradation, 2) loss of situational awareness, 3) automation brittleness and literalism, and 4) increased time and difficulty to diagnose failures and manually take over when required [3, 29]. If the LOA is too low, potential HSC problems can include: 1) cognitive and working memory overload in routine tasks under time pressure, 2) human decision biases and heuristics, 3) lack of repeatability and consistency, 4) complacency and boredom, and 5) greater human interdependency and chaos under failure [3, 29]. As a consequence, care must be taken to consider each of the roles human and machine should perform in a given task, and automation only introduced when there is a specific need to do so.

2.2.2 The Impact of Automation on Human Performance

Several previous studies have investigated levels of automation in the context of multiple UAV supervisory performance. Ruff et al. [35] conducted a study that looked at the effects of level of automation and decision-aid fidelity in human interaction with 1,
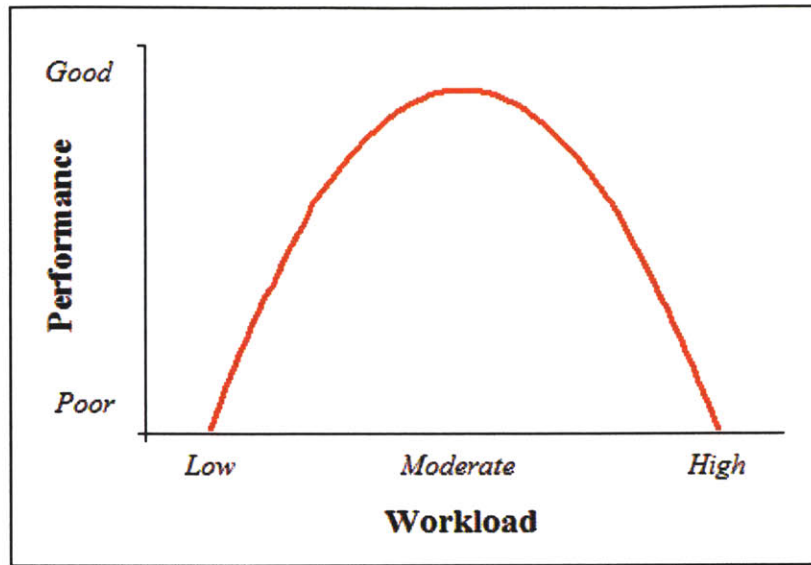
Figure 2-1: The Yerkes-Dodson Law [46]

2 and 4 UAVs. They found that a medium level of automation called management-by-consent, which corresponds to an automation level of 5 on the scale of Table 2.1, had significant advantages over manual control (Level 1, Table 2.1) and management-by-exception (Level 6, Table 2.1) supervisory control schemes. This level of automation had the highest levels of operator situation awareness (SA) and performance, though subjective workload ratings were not always better. Performance declined sharply as the number of supervised UAVs increased in the manual control condition, suggesting significant levels of automation are required for acceptable performance in supervising multiple vehicles.

Ruff et al. [34] later examined the effects of automation on task completion time and subjective workload levels in the control of 2 or 4 UAVs. A management-by-consent level of automation (Level 5, Table 2.1) was compared to management-by-exception (Level 6, Table 2.1), and no significant differences in performance or workload were initially found between the levels. It was hypothesized this result was due to a common behavior exhibited by subjects where they were likely to act on the automation's recommendation before the automation carried out the action under LOA 6. Subjective workload ratings were higher and performance lower for 4 UAVs

than with 2. A follow-up experiment then kept the number of UAVs constant at 4 and the time limit to over-ride the management-by-exception automation was made an independent variable. It was found that there still was no performance difference between automation conditions, but shorter time limits to over-ride the computer in management-by-exception caused higher workload and poorer performance. An important finding from this study was that subjects using management-by-exception (Level 6, Table 2.1) did not use the automation as intended. Upon receiving a notification of what the automation was planning to execute, users consistently executed it themselves manually instead of letting the computer take over. This suggests that subjects may not have been properly evaluating or did not trust the computer's recommendations.

Under the Defense Advanced Research Projects Agency (DARPA) Mixed-Initiative Control of Automata (MICA) program, some research has been conducted into the development of unmanned combat aerial vehicle controllers that enable small teams of operators to effectively task large teams of aircraft. Linegang et al. [22] set out to determine what information requirements human operators needed to effectively use and interact with such systems. They found that it was particularly difficult to convey to the user a dynamic understanding of a mission in both space and time, which is critical for high levels of operator SA. Timeline and map-based components of an interface were recommended to overcome this. Under the same research program, Roth et al. [32] evaluated a prototype mixed-initiative interface intended for the supervisory control of multiple vehicles, with the intent of investigating to what extent human operators were able to understand mission plans for the UAVs that were generated by automated controllers. They found that while humans could understand the plans created, they could not understand the rationale behind them. Therefore, operators had low SA as they were not able to evaluate the effectiveness of the automatically generated plans and could not predict the effects of changing them. The conclusion that can be drawn is that in multiple UAV control, operators using high levels of automation may perform very well under nominal conditions but likely will not be able to cope as well with uncertainty in the mission plan.

Wilson and Russell [45] investigated whether adaptive automation based on psychophysiologic cueing could be used to relieve operator taskload during periods of high workload, and in so doing improve task performance. Subjects performing a UCAV target identification task showed improved hit rates on targets and less missed weapons release points under adaptive aiding, thus showing the benefits of mitigating operator workload in potential overload situations. However, the adaptive aiding used in the study decreased the velocity of the UAV and did not actually change the level of automation for the human controller. Modifying mission parameters to relieve operator workload is often not a viable option in time-critical applications, as UAVs generally must operate within strict windows of time. This is an example of what can occur when system designers only incorporate human performance considerations into their designs and ignore hard environmental (system) constraints.

2.3 System Performance

2.3.1 Wait Times

In previous work modeling human-robot (ground-based) interaction and operator capacity for supervisory control, it has been proposed that the number of robots or vehicles a single individual can control is given by Equation 2.1 [17, 26, 27]. In this equation, FO (Fan Out) equals the number of robots a human can effectively control, NT (Neglect Time) is the expected amount of time that a robot can be ignored before its performance drops below some acceptable threshold, and IT (Interaction Time) is the time it takes for a human to interact with the robot to ensure it is still working towards mission accomplishment. The addition of one in Equation 2.1 represents the baseline condition in that an operator can control a single robot. While originally intended for ground-based robots, this work has direct relevance to more general human supervisory control tasks where operators are attempting to simultaneously manage multiple entities, such as in the case of UAVs. Fan out is a measure of workload, as the smaller the ratio of NT to IT gets, the greater the proportion of

time that the operator has to spend attending to a single vehicle.

$$FO = \frac{NT}{IT} + 1 \qquad (2.1)$$

Modeling interaction and neglect times are critical for understanding human workload in terms of overall management capacity, but there remains an additional critical variable that must be considered when modeling human control of multiple robots, regardless of whether they are on the ground or in the air, and that is the concept of Wait Time (WT). In HSC tasks, humans are serial processors in that they can only solve a single problem or task at a time [4], and while they can rapidly switch between cognitive tasks, any sequence of tasks requiring complex cognition will form a queue and consequently wait times will build. In the context of a system of multiple vehicles or robots in which two or more vehicles will likely require attention simultaneously from a human operator, wait times are significant in that as they increase, the actual number of vehicles that can be effectively controlled decreases, with potential negative consequences on overall mission success. Figure 2-2 illustrates how wait times could impact an overall system.

In multiple vehicle supervisory control, operators interact with a robot or vehicle to bring its performance to some acceptable performance threshold and then neglect it until such time that it requires assistance. Performance may degrade gradually over time (NT_2 in Figure 2-2), or very suddenly with a discrete event (NT_1 in Figure 2-2). For example, if a robot has a directional gyro that is slightly miscalibrated and it is instructed to autonomously reach a certain navigational point, its position error increases gradually over time until the human chooses to interact with it again. A discrete event causing a sudden drop in performance such that a robot requires immediate operator assistance is a system failure or the need for clarification of a goal state. Sliding degradation is much harder for operators to correct, because the rate of change of a system state may be so gradual that the operator does not detect it immediately. In this case, performance is significantly degraded without the operator realizing that they need to intervene. In either case, as soon as the robot's

35

performance level decreases below a specified level or threshold, wait time begins. This is shown as a static, constant line on Figure 2-2, but this performance threshold may change depending on the task at hand.

While interaction and neglect times are important in predicting human capabilities for handling multiple robots, for those domains that are time-critical and high risk like UAVs, WT becomes a critical point for possible system failure. In many ground-based robot applications such as mine-sweeping, waiting for human interaction may not be critical, but certainly for UAVs and UUVs (unmanned underwater vehicles) with expected time on targets and dynamic threat areas, waiting is not only sub-optimal, it can be extremely hazardous. While most robots and vehicles can be programmed to follow some predetermined contingency plan if they do not receive required attention, mission success will likely be significantly degraded if wait times grow unexpectedly.
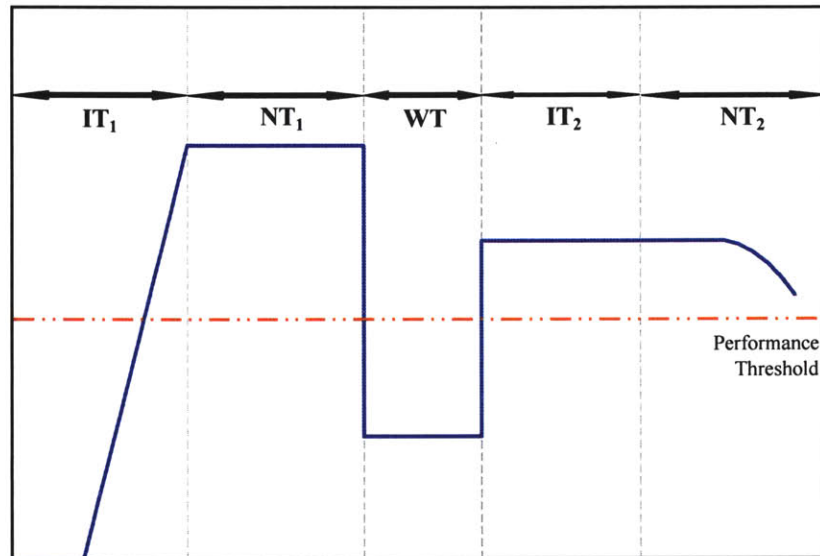
Figure 2-2: The Relationship Between IT, NT, and WT

From the robot or vehicle perspective, WT imposed by human interaction (or lack thereof) can be decomposed into three basic components: 1) wait time in the human decision-making queue (WTQ), 2) interaction wait time (WTI), and 3) wait time due to loss of situation awareness (WTSA). For example, suppose an operator is control-

ling two robots on a semi-autonomous navigation task (much like the Mars Rovers). While typical operations involve human interaction with a single vehicle, there will be times when both vehicles require attention simultaneously or near-simultaneously. When this occurs, if the human operator begins assisting the first robot immediately, the first robot must wait while the operator solves the problem and then issues commands to it (WTI_1). For the second robot, the time it waits in the queue (WTQ_2) is effectively WTI_1. If an operator does not realize a robot or vehicle needs attention, the time from the initial onset of the event to actual operator intervention could range from seconds to minutes. This wait time induced by lack of recognition for required intervention is an example of WTSA.

IT, which was previously defined as the time it takes for a human to interact with a robot, can be further decomposed. IT is the time during which a human's attention is focused on a single robot in order to solve a problem or induce some change to improve performance above a specified threshold. From the human perspective, IT includes the time required to determine the nature of the problem, solve the problem, and communicate that solution to the robot, with some type of feedback. Thus the robot must wait some period of time during the "interaction" due to the human decision-making process. In teleoperation where the human is directly controlling a robot's movements and positions, interaction wait times might be very small, and occur in rapid succession as the controller adjusts commands according to sensor feedback. However, in other scenarios that require minimal manual control but significant cognitive input such as the need to provide a new mission to a UAV, WTI can be quite large depending on the complexity of the problem. Previous research has indicated that system interaction times of operators should not exceed 0.7 of the total system operating time due to cognitive and physical limitations [33, 36].

$$WT = \sum_{i=1}^{X} WTI_i + \sum_{j=1}^{Y} WTQ_j + \sum_{k=1}^{Z} WTSA_k \qquad (2.2)$$

$$FO = \frac{NT}{IT + WT - \sum_{i=1}^{X} WTI_i} + 1 \qquad (2.3)$$

37

Equation 2.2 defines the definition of wait time to be used for the remainder of this thesis. It categorizes total system wait time as the sum of the interaction wait times, which are the portions of IT that occur while the vehicle is in a degraded state (WTI), wait times that result from queues due to near-simultaneous arrival of problems (WTQ), plus wait times due to operator loss of SA (WTSA). In Equation 2.2, X indicates the number of times an operator interacted with a vehicle while the vehicle was in a degraded state, Y indicates the number of interaction queues that build, and Z indicates the number of time periods in which a loss of SA causes a wait time. Equation 2.3 demonstrates how the Fan Out equation (Equation 2.1) would change as a result of the inclusion of wait times. WTI must be subtracted from the denominator of Equation 2.3 because WTI is a subset of IT, and WTI's inclusion in the wait time formula means it would otherwise be replicated. The addition of wait time to the denominator of this equation is likely to significantly decrease the predicted number of vehicles one person can control, as aggregate wait time can be many times the size of IT.

2.3.2   The Impact of Automation on System Performance

It is not clear what effect the level of automation of a human supervisory control task has on the various types of wait times, and what the overall system effectiveness will be as these wait time components change in magnitude and proportion to one another. The following section will make predictions of trends across different levels of automation which will be compared to results later on in this thesis.

Interaction Wait Time (WTI)

Olsen and Goodrich [26] described four components of IT as 1) vehicle monitoring and selection (VMS), 2) context switching and acquisition (CSA), 3) planning or problem solving (PPS), and 4) command expression (CE). As outlined above, interaction wait time is a subset of IT that occurs when any or all parts of these stages take place while the vehicle requires human input. In general, increasing levels of decision support

should decrease vehicle selection time, as a primary goal of automation in this context would be to more quickly identify problem states. Context switching and acquisition, the next step in the interaction process, occurs when operators attempt to update their knowledge of the new vehicle's current goals and problems. The process of CSA in conjunction with VMS can incur a "switching cost" in which switching between tasks incurs added cost in terms of wait time because of the cognitive need to orient to the new problem. Switching costs are not incurred simply as a function of change detection, but occur as an operator regains the correct mental model and situation awareness needed to solve the new problem.

Switching costs in terms of added wait times will occur because in the control of multiple UAVs, operators spend time monitoring unfolding events, but periodically engage in interactive UAV control tasks. This need to concentrate on a task, yet maintain a level of attention for alerts causes operators to have a conflict in mental information processing. Interrupt-driven processing, needed for monitoring alerts, occurs when people are sensitized to possible problems and expect distraction. This is the mode operators supervising multiple UAVs will nominally find themselves in when missions are executed according to plan. Concentration on a task, like that needed for UAV intervention, requires task-driven processing which is likely to cause decreased sensitivity or attention to external events. While interrupt and task driven processing can both be present in a person, attention must be shared between the two and switching can incur cognitive costs that can potentially result in errors [24]. Therefore, switching costs are expected to be higher for very low levels of automation as there is a greater demand for task-driven processing at low automation levels.

The planning/problem solving stage of interaction time occurs when the operator plans a course of action for the selected vehicle. In general, increasing levels of automation should lower planning times because the computer takes progressively more decision options away from the human and/or aids the human by executing some of the planning steps. As the human is presented with fewer or no alternatives, they have a smaller problem space to explore, though this limitation may result in less than ideal solutions.

The last stage of interaction wait time is the execution or command expression stage. In this stage, users must express their intent to the vehicle by mapping it onto a series of actions that the vehicle can understand. Increasing levels of automation are likely to decrease this time as well because higher levels of automated decision support may not require human command intent to execute some or all functions, or it may offer the human a single or limited options to choose from in terms that the vehicle already understands.

With the exception of CSA, all the components of WTI follow the same trend of decreasing with increasing levels of automation. Therefore, unless CSA dominates all other components of WTI at very high or very low levels of automation, WTI should decrease with increasing levels of automation.

Wait Time in the Queue (WTQ)

In the context of human supervisory control and multi-UAV management, the elements of an operator's queue are tasks that the operator must perform, such as firing on a target, re-planning a route, or assigning an emergent target to a particular UAV's mission plan. In the context of queuing theory, the operator's tasking can be thought of as a preemptive priority queue with a single server. The time of the service rate, the average time an operator takes to attend to a vehicle, is essentially the average WTI. As previously mentioned, WTI for the vehicle in service corresponds to additional queuing wait time for all the vehicles in line. From the above discussion, WTI is predicted to decrease with increasing levels of automation, so the same trend can be expected with service rate. The arrival rate is the average time between tasks that the operator must perform, and is dependent upon the scenario complexity and the number of UAVs to be controlled, amongst other things. Within this framework, it can then be seen that utilization of the operator will decrease with increasing level of automation, and therefore the average wait time in the queue will decrease, potentially in a non-linear way. A small increase/decrease in individual vehicle WTI can have a much larger impact on WTQ, particularly as the number of tasks in the queue becomes large. The implications of this are that there could be a much steeper

40

increase in WTQ at the lowest levels of automation.

Situation Awareness Wait Time (WTSA)

WTSA is perhaps the most difficult wait time component to model because it represents how cognitively engaged and aware an operator is in the task. Situation awareness is generally defined as having three levels, which are: 1) the perception of the elements in the environment, 2) the comprehension of the current situation, and 3) the projection of future status [13, 14]. An example of a WTSA would be the failure of an operator to notice a message from a UAV that notified her of a failure which rendered it useless for the remainder of a mission, such as an inability to release weapons. The time it takes for the operator to process the message and task the appropriate UAV to return to base would be a WTSA. While notifications and critiquing devices included in decision support systems can help to alleviate added wait time due to loss of SA, it is still an event that at the very least, should be included as a probabilistic variable in a larger model of wait time for human interaction with multiple vehicles.

As an operator's level of SA can decrease under high workload due to competition for attentional resources [43], but also decrease under low workload due to boredom and complacency [31], it can be concluded that optimum level of operator SA occur under moderate levels of workload. It is predicted that WTSA will follow the opposite trend. Therefore, medium levels of automation should have the lowest accumulated WTSA, while very low or very high LOAs should have higher WTSA.

Very high levels of automation should eliminate any wait time due to the loss of SA, but only for planned events. A primary concern with highly automated systems is that when an unanticipated event occurs, automated systems often arrive at erroneous solutions and humans do not have enough SA to recognize the failure mode. Because of the propensity of human towards automation bias in command and control settings [5] which can be exacerbated by the loss of SA, it is possible that an operator will not veto erroneous automated actions, thus causing some potentially catastrophic event. Indeed, this problem was seen recently in the 2004 war with Iraq when the

U.S. Army's Patriot missile system engaged in fratricide, shooting down a British Tornado and an American F/A-18, killing three. This avoidable loss of life occurred because human operators did not recognize the guidance system had erroneously locked onto aircraft instead of enemy missiles. It is exactly in this kind of instance where highly automating the system to reduce wait times caused by humans should be very carefully considered.

2.3.3 Wait Time Costs

The preceding discussion on wait times has not mentioned their cost. A wait time cost measure is needed because wait times alone don't quantify how much impact they have on a particular mission. It can be inferred that higher wait times are related to higher costs in performance, but the threshold at which they incur a cost needs to be identified. In addition, the context in which wait times occur may have an even greater influence on overall system performance. For example, a typical mission for a UCAV such as a Predator involves striking targets at precise times. If the mission planner has built in additional "slack" time into a route so that the UAV may incur wait times without missing its deadline, then the cost of wait time is relatively low until all of the slack time is used, whereas any additional wait time will cause the UAV to fail to destroy a target on time. This could have a very high potential cost if the target was of strategic importance. Freed et al. [15] has quantified wait time cost with respect to UAV surveillance for the purpose of evaluating computer algorithms, but more research is needed in this area.

2.4 Conclusion

Several studies, particularly those conducted by Ruff et al. [34, 35], have investigated what levels of automated decision support are appropriate for multiple UAV supervisory tasks. The general conclusions are that manual levels of decision support might perform adequately when controlling smaller numbers of UAVs, but larger numbers of UAVs quickly overwhelm operators. Higher levels of decision support, LOA 6 and

higher from Table 2.1, appear to have performance advantages over lower levels of decision support under nominal conditions, but operators are likely to have degraded SA and therefore cannot adapt to abnormal, unexpected conditions. Review of current literature suggests that to strike the best balance between performance and SA, moderate levels of automation, LOAs 4 or 5 (Table 2.1) should be used.

However, all of these studies focused solely on human performance metrics and did not consider the overall system in their evaluations. From the system performance perspective, it is important to model the sources of wait times, especially since these times could potentially lead to system failure. The impact of increasing levels of automation on overall wait times is unclear, but some trends can be predicted for individual wait time components. In particular, wait times due to interaction and queuing should decrease with level of automation, while those due to situation awareness should increase only at very low or high levels of automation.

The remainder of this thesis will seek to extend previous work on appropriate levels of automation for time-critical multiple HSC tasks, specifically for UAV supervision, but from a mission-centered as opposed to a human-centered perspective. In addition to the previous studies that examined automation strategies, in this research, human and system performance will be measured simultaneously and evaluated both individually and against each other.

Chapter 3

Simulation and Interface Design

This chapter presents the details of a multiple vehicle supervisory control simulation and interface developed to study the effects of automation levels on human and system performance in time critical applications. Explanations of this program's functionality, appearance, and usage are outlined, and the rationale behind the program's design is offered.

3.1 Overview

In order to study how levels of automation affect temporal constraints in human supervisory control multiple task management, a dual screen simulation test bed named the Multi-Aerial Unmanned Vehicle Experiment (MAUVE) interface was developed (Figure 3-1) that allows an operator to effectively supervise four UAVs simultaneously, and intervene as the situation requires it. Using this interface for subsequent experiments, subjects took on the role of an operator responsible for supervising 4 UAVs collectively tasked with destroying a set of time-sensitive targets in a suppression of enemy air defenses (SEAD) mission. The simulated UAVs were highly autonomous, and therefore only required operators to provide high level mission planning and input execution actions to the UAVs. The UAVs launched with a pre-determined mission plan that came from an air tasking order (ATO), so initial mission planning to include target assignments and routes was already completed prior to launch. The
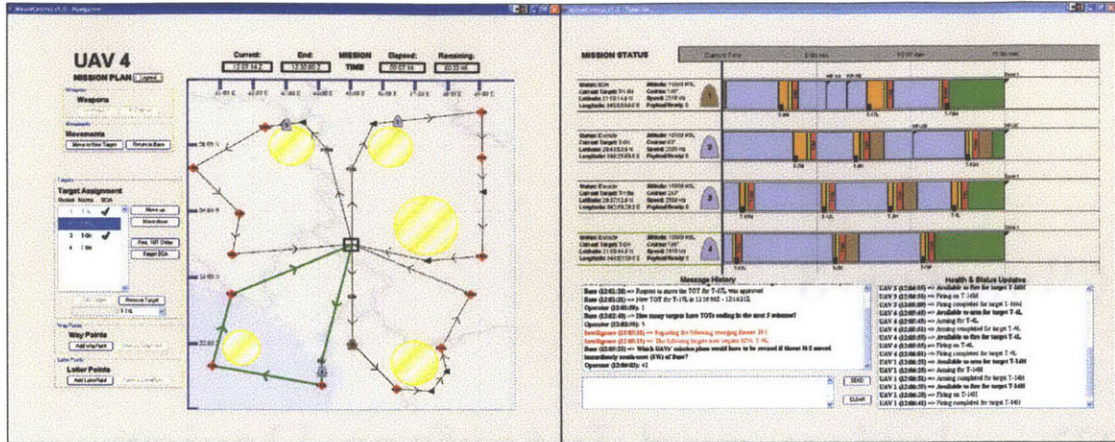
Figure 3-1: The MAUVE Dual Screen Interface

operator's specific tasking in the MAUVE simulation was to monitor each UAV's progress, re-plan aspects of the mission in reaction to unexpected events, and in some cases manually execute mission critical actions such as arming and firing of payloads.

3.2 Navigation Display

The UAVs supervised by subjects in MAUVE were capable of 6 high-level types of actions in the simulation: 1) traveling enroute to targets, 2) loitering at specific locations, 3) arming payloads, 4) firing payloads, 5) performing battle damage assessment, and 6) returning to base. Battle damage assessment (otherwise known as battle damage imagery or BDI) is the post-firing phase of weapons release where it is determined whether the weapon(s) hit the target, and if the desired effect was achieved. Table

UAV Action	Color
Enroute	Blue
Loitering	Orange
Arming Payload	Yellow
Firing Payload	Red
Battle Damage Assessment	Brown
Return to Base	Green

Table 3.1: Color Coding of UAV Actions in MAUVE

46
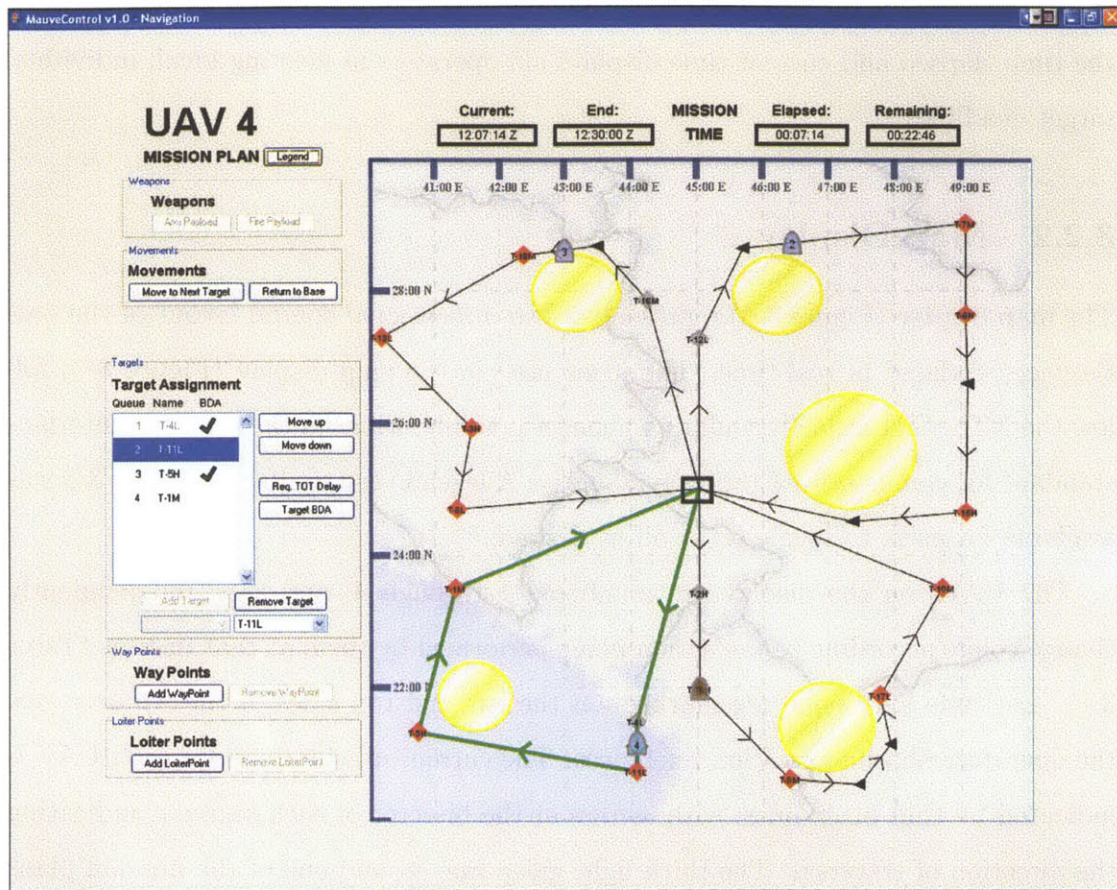
Figure 3-2: The MAUVE Navigation, Mission Planning and Execution Display

3.1 outlines the color coding assigned to each of these actions in the simulation.

The left-hand side of the MAUVE interface is known as the navigation display, and it consists of a mission time window, map display, and a mission planning and execution bar (Figure 3-2).

3.2.1 Mission Time Window

At the top right of the map display is a mission time box showing both time elapsed and time remaining in absolute and relative terms. Due to the time critical nature of the set of targets to be destroyed and the large number of tasks to be performed while supervising four UAVs, time management and scheduling is the fundamental problem faced by operators in MAUVE. Operators are supported in meeting the

47

global deadline for the mission by the time remaining and end of strike timers, while the time elapsed and current time displays aid operators in meeting local, individual target deadlines.

3.2.2 Map Display

The map display (Figure 3-3) represents a two-dimensional spatial layout of the battlespace, updated in real-time. The icons used in the map display (Figure 3-3) follow the MIL-STD-2525B[1]standard whenever possible, which is the US DoD interface standard for common warfighting symbology. A legend very similar to Figure 3-3 was available for users to toggle on or off as desired.

The UAVs on the display are numbered 1 through 4, and they independently change colors according to the action being performed by them at that instant (Table 3.1). Latitude and longitude markers on the left and top sides of the window give the operators absolute location references. The current mission plan for each UAV is indicated by thin black lines, with arrows at the bisector of each segment indicating the direction of traversal. The thick light green line around one of the mission plans indicated that plan was the one currently selected by the user. For instance, in Figure 3-3, UAV4's route has been selected by the user.

Targets are designated by a diamond-shaped icon, and are assigned a relative importance to the mission plan (priority) of high (H), medium (M), or low (L). All target names follow a naming convention of T-XXP, where "T" indicates it is a target, "XX" is a unique identifying number, and "P" is its priority. All target names are super-imposed on the relevant icon (Figure 3-4) and active targets are differentiated from inactive targets by their color, which is either red or gray on the display, respectively. An inactive target is any target that had either was destroyed or its TOT deadline passed.

Waypoints, which are shown on the map display with black triangle icons, represent UAV turn points while enroute. Naming of waypoints follows the convention

---

[1]A military standard (MIL-STD) is a specification which lists and explains a compilation of prerequisites that an item must meet for US DoD acceptance

(a) Legend   (b) Map Display
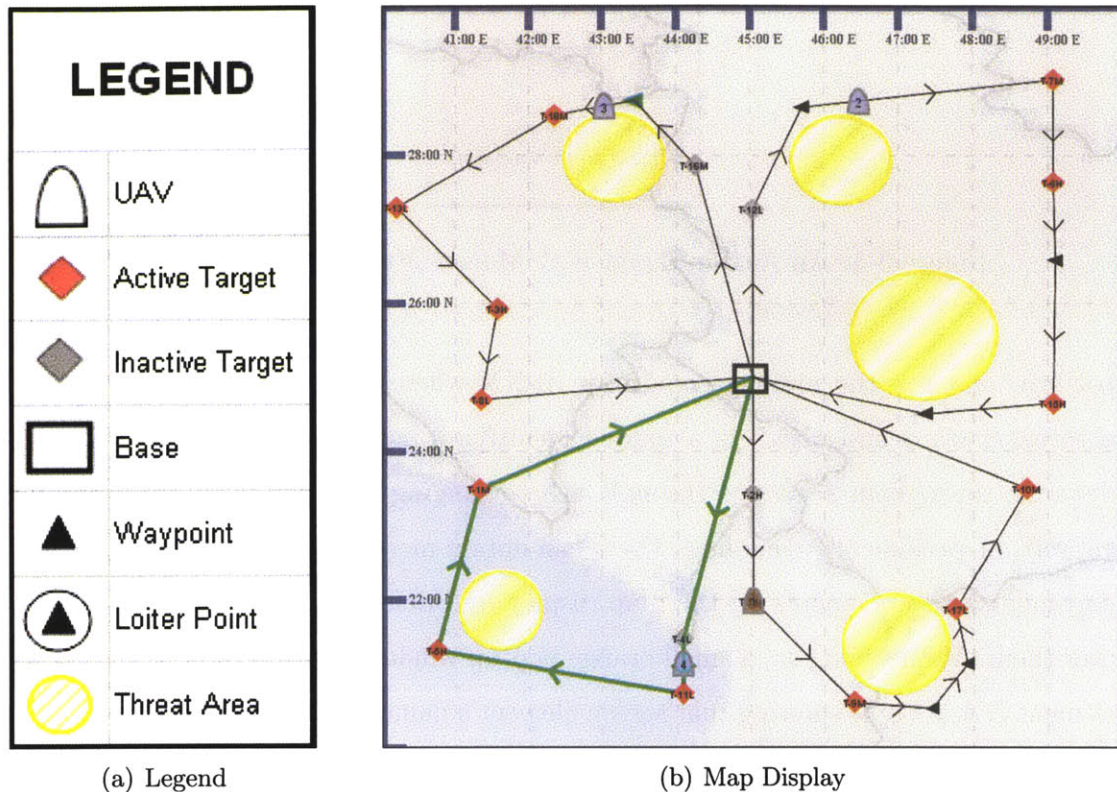
Figure 3-3: The Map Display with Legend

WP-XY, where "WP" indicates the object is a waypoint, "X" is the UAV number that the waypoint is associated with, and "Y" is a unique identifying letter for the specific route. Loiter points are represented in a similar fashion. The naming conventions are identical, except loiter point names begin with the letters "LP" and their icons have an additional circle around the black triangle graphic. Functionally, a loiter point is the same as a waypoint except that when a UAV reaches a loiter point, the UAV loiters for a user-specified amount of time before moving on. Minor adjustment to UAV routes on the map display can be made by selecting a particular waypoint or loiter point (indicated by a dark green highlighting as seen at the top of Figure 3-3) and dragging it across the display to the desired location. More significant routing changes such as the addition or removal of waypoints, loiter points, or targets (together known as navigation points) is accomplished using the mission planning and execution bar which is described in Section 3.2.3.

Threat or hazard areas are always circular in shape with a striped yellow coloring

49
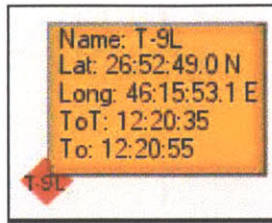
Figure 3-4: An Active Target with Mouse-Over Pop Up Window

pattern, and use the naming convention H-XX where "H" indicates the object is a threat/hazard and "XX" is a unique identifying number. Threat areas can be dynamic throughout scenarios through either changing size, locations, disappearing entirely, or periodically emerging. Users can obtain more detailed information about particular screen elements on the map display by briefly hovering the mouse cursor over them. While doing so, a small orange pop-up window appears next to the screen element (Figure 3-4) showing that screen element's name, position in degrees latitude and longitude, and if a target, the beginning and end of its TOT window. The TOT window start and end were designated by the ToT and To headings, respectively (Figure 3-4).

3.2.3   Mission Planning and Execution Bar

Located on the far left of the navigation display, the mission planning and execution bar (Figure 3-5), supported the majority of human interaction with the UAVs. Each UAV has its own mission planning and execution bar that is selected by either clicking on the desired UAV's status window on the decision support display (Figure 3-7), or on the UAV icon in the map display. As described in section 3.2.2, light green highlighting around the UAV's status bar and its current mission plan in the map display told subjects which UAV and route was currently selected.

Mission Planning

Through the mission planning interface, users are able to manipulate the set of targets each UAV visits (including order of visitation), their deadlines, and the path the UAV

travels between them. A key element in mission planning is the target assignment queue window (Figure 3-5). The target assignment queue lists all targets currently assigned to a UAV in the order they were to or did visit, so it serves as a time history and future prediction of all targets currently assigned to that UAV. Just like the map display (Section 3.2.2), inactive targets are grayed out in the list to differentiate them from active targets. Battle damage assessment is toggled to the opposite setting of "on" or "off" for a particular target by selecting it in the target assignment queue and clicking the "Target BDA" button to the immediate bottom right of the window. A check mark beside a target indicates that battle damage assessment is currently "on", or scheduled for that target. BDA is semi-automated in the sense that the operator is responsible for scheduling BDA in advance, but the UAV performs it automatically after firing, if scheduled. Due to the planned approach to this function, the user chooses whether BDA is performed on a target before the arming window occurs. Subjects can also change the order that targets are visited by hitting the "Move up" or "Move down" buttons beside the top right of the target assignment queue. In doing so, the mission paths on the map display are automatically re-planned with straight line paths between the new set (or ordering) of targets. All existing waypoints and loiter points on the affected paths are deleted.

The "Request TOT Delay" button allows users to have a limited opportunities to manipulate the set of TOTs they are attempting to meet in any given scenario. Subjects can request a time-on-target (TOT) delay for a given target for two reasons: 1) According to the current mission plan, a UAV is predicted to arrive late to that target and therefore miss its deadline, or 2) for workload purposes, i.e., if a subject feels they need to spread out workload to manage the UAVs more effectively. However, this function must be used with care because moving back one target's deadline affects a UAV's arrival time at all subsequent targets. Users have to be careful that they are not causing additional missed targets farther in the future in return for near term gains. This change of TOT is a request, not a command, and the request can be approved or denied. The probability of approval is a function of how far in advance of the deadline the request is sent, as would likely be the case in true military situations.

The probability distribution for chance of approval is given by Equation 3.1. Subjects do not know the actual probability function but are aware that if they want to move a deadline, they have a greater chance of doing so farther in advance.

$$P(Approval) = 1.0 - e^{-t/450} \tag{3.1}$$

where t = time in seconds before deadline request was sent

When a TOT deadline is immediately approaching, the chance of approval is near zero, but nearly 1.0 when requested 15 minutes in advance, which is as far ahead as the decision support shows for higher levels of automation. A request always takes 5 seconds to come back, and during this intervening time no other TOT requests for any other targets can be made. Users can request as many TOT delays as they wish for a given target, but there is no guarantee of ever having one approved. Requesting TOT delays requires focused operator attention for extended periods of time, which can be problematic because it may be more appropriate to focus on another more time-critical task, such as re-planning a route due to an emergent threat.

Targets can be added and removed from specific UAV's routes by hitting the "Add Target" and "Remove Target" buttons. To add a target, subjects first select from the list of eligible targets in the pull-down box immediately below the button. In order for a target to be eligible to be added to a route, it cannot be a part of any other route. If a user wishes to switch a target to a different route, they must first remove it from one route before adding it to another. This was done deliberately to ensure a target could only be assigned to a single UAV. Frequently there are no unassigned targets in the battlespace, so in this case the option to add a target was disabled entirely, as seen in Figure 3-5. The "Remove Target" button works similarly, the only difference is that the target to be removed can additionally be selected by clicking on it in the target assignment window. The "Remove Target" button is primarily used to remove particular targets from the mission plan after a re-planning message gives orders to do so, but it is also used to re-assign targets between different UAVs. The "Add Target" function is required when an emergent target appears, when a missed target

52
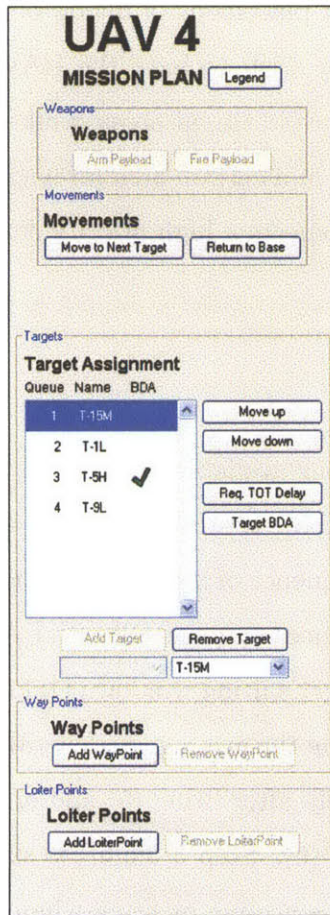
Figure 3-5: The Mission Planning and Execution Bar

is added back to the schedule, or when an operator wants to assign a target to a different route.

The final mission planning functions available to users of the MAUVE interface are the ability to add and remove waypoints and loiter points. Removing waypoints and loiter points is a simple 2 step process. First, users click on the waypoint or loiter point on the map display that they want to remove, and then they hit the "Remove Waypoint" or "Remove LoiterPoint" button. To add a waypoint, users hit the "Add Waypoint" button and click on the map display where they want to place it. Based on the currently selected route, the simulation attempts to add it to the nearest route segment. From there, users drag the waypoint to any desired location on the map display. Adding a loiter point through the "Add LoiterPoint" button

worked similarly, except upon placement, a dialog box appears on the map display that asks users how long they wish to have the UAV loiter at that specific point. Users could add waypoints to avoid threat areas, while loiter points could be used if a user wanted a UAV to remain in a certain area for a specific reason. One such reason would be that the operator suspects a high priority emergent target might appear at a later time.

Mission Execution

The mission execution functions available in MAUVE are arming, firing, performing BDA (described in Section 3.2.3), returning to base, and moving to another target. Figure 3-6 shows a nominal sequence of mission execution events that a UAV performs at an active target. The sequence occurs as follows: UAV 4 arrives at active target T-1L before the start of that target's firing window (1), loiters until the beginning of the arming window (2), arms during the arming window (3), fires during the TOT window (4), performs BDA immediately after firing (5) and after finishing BDA, moves onto the next target automatically (6). Step 2 might be skipped if the UAV arrived late to a target, i.e., in that target's arming or firing windows. Step 6 (performing BDA) was not required at every target.

The "Arm Payload" and "Fire Payload" buttons on the Mission Planning and Execution Bar (Figure 3-5) are only enabled if the rules of engagement (RoE) (Appendix A) of the simulation are met. For arming, this means that the UAV is directly on top of a target and the mission time is within the arming or firing windows, and for firing it means that the UAV is at the relevant target, within the arming window, and already armed. In either case, the arming or firing buttons are not enabled unless the relevant action is predicted to finish by the end of the firing window.

The "Move to Next Target" button causes the selected UAV to start moving enroute to the next scheduled target, and can only be used in off-nominal situations. It is required because the automation rules governing UAV behavior cause every UAV arriving at an active target to loiter indefinitely unless the user intervenes, such as arming and firing on that target. However, if arming and firing does not occur within

54
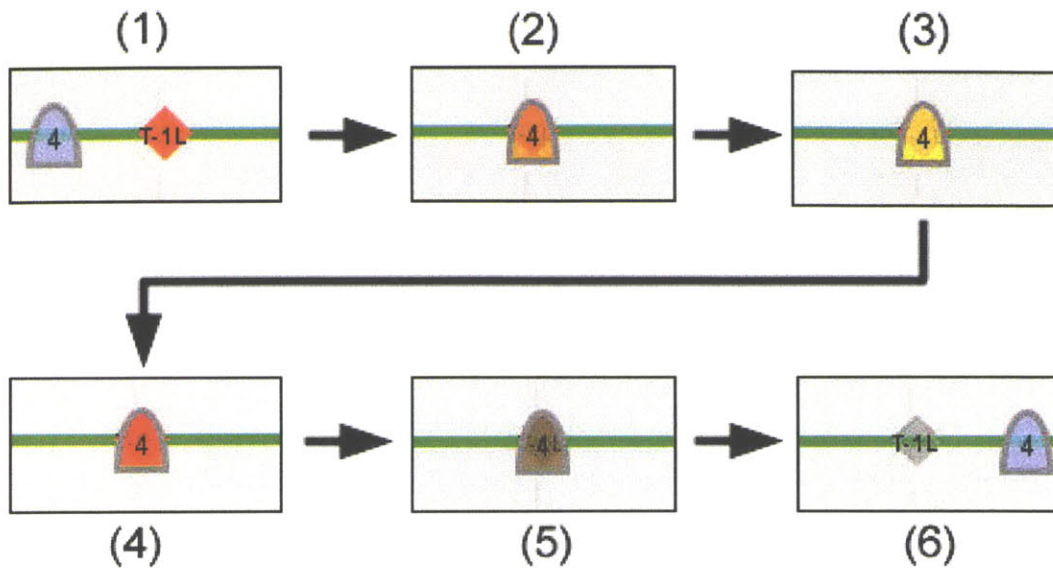
Figure 3-6: A Series of Typical UAV Actions at an Active Target

the appropriate windows, then that target is missed and the user needs a way to get the UAV moving again. There could be situations in the simulation where the RoE changes such that a previously designated target should not be destroyed, thus in this case arming and firing on certain targets is not desired. In terms of Figure 3-6, the "Move to Next Target" button is a way to go from Step 2 to Step 6 directly.

The "Return to Base" function causes all future targets, waypoints and loiter points to be deleted from the individual UAV's mission plan, and a straight line path for that UAV is planned directly back to base. It is used whenever subjects want to get a UAV back to base as soon as possible.

3.3 Decision Support Display

The right-hand side of the MAUVE simulation is termed the decision support interface, and it consists of a UAV status window, chat box, UAV health and status updates, and the decision support window (Figure 3-7).
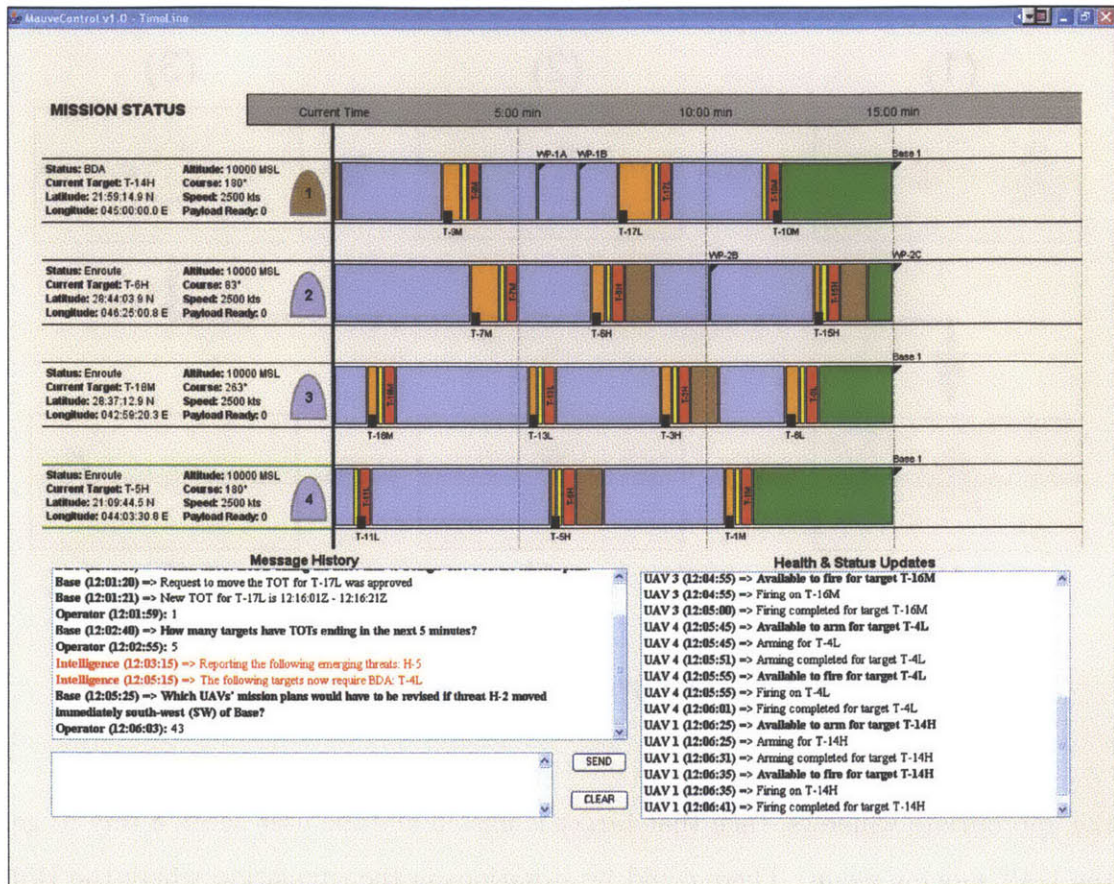
Figure 3-7: The MAUVE Decision Support Display

3.3.1 UAV Status

The top left of the decision support display contains a status window (Figure 3-8) that gives the operator low level, detailed information for each UAV. This window shows each UAV's current target, current status, position in latitude and longitude, course, and weapons information. The current target field indicates which target the UAV is currently scheduled to reach or the target it was currently loitering at, depending on whether the UAV was enroute or loitering. The UAV's current action field has a description consistent with the color of the UAV on the map display (Figure 3-3) as well as the icon on the right of the panel. The payload ready indicator always reads either "1" or "0". If it shows "1" instead of "0", this means that the UAV has armed for its current target, and is available to fire.

56

Figure 3-8: The UAV Status Displays

Speed and altitude are also shown in the status display, although they are not directly controllable by operators. Each simulation is run approximately 4 times faster than real time so an entire strike can take place over 30 minutes (instead of several hours as is commonplace in real life strikes), thus the speeds shown are not representative of true UAV simulated velocities.

3.3.2 Chat Box

The bottom left of the decision support display holds a text-based communication tool known as a chat box (Figure 3-9) that contains a time history of all human interaction. The chat box is an established method of communications in current day military command and control scenarios [6]. The use of a chat tool in the MAUVE simulation adds ecologic and external validity as a secondary workload and situation awareness tool [8].

The chat box window displays various notification messages that appear in re-

Message History

Base (12:07:18) => New TOT for T-5H is 12:17:10Z - 12:17:30Z
Base (12:07:40) => Which future high priority target would provide the greatest decrease in a single UAV's travel time, if removed from the ATO?
Base (12:07:47) => Request to move the TOT for T-9L was approved
Base (12:07:47) => New TOT for T-9L is 12:21:06Z - 12:21:26Z
Operator (12:08:30): 10
Intelligence (12:08:45) => Reporting the following emerging threats: H-4
Base (12:08:55) => How many active medium priority targets remain to be destroyed?
Intelligence (12:09:30) => Reporting the following emerging targets: T-18M
Intelligence (12:09:40) => The following targets have been re-assigned to another strike mission: T-14L

SEND

CLEAR

Figure 3-9: The Chat Box Window

sponse to scenario events or actions taken by users, as well as periodic task-relevant questions for operators to answer. A scroll bar on the right side of the chat box ensures that subjects can scroll up to see a complete time history of all messages for the current session. The accuracy and time delay in responses to the online queries are measured to obtain an objective measurement of situation awareness as well as secondary workload, or spare capacity. Messages in red are from intelligence commanders and always require an immediate re-planning response by the operator. An arrival of one of these high priority messages is signaled by an audio alert to ensure operators were aware of these events as they occurred. Messages in bold black are questions about the current situation from a remotely located commander or superior officer. A response is required, but is a lower priority task than supervising the UAVs. Messages in plain black are purely informational messages that do not require a response. One informational message that is particularly important to operators are notifications that TOT requests had been accepted or denied. If a TOT request is denied, a pop-up box appears on top of the map display, an audio alert sounds, and a message stating the denial appears in the chat box. If it is approved, the requested target's TOT is moved and a message stating the approval and new deadline for that target appears in the chat box.

58

Health & Status Updates
UAV 2 (12:15:19) => Arming for T-11M
UAV 2 (12:15:25) => Arming completed for target T-11M
UAV 2 (12:15:26) => Available to fire for target T-11M
UAV 2 (12:15:27) => Firing on T-11M
UAV 2 (12:15:33) => Firing completed for target T-11M
UAV 1 (12:21:26) => Available to arm for target T-13L
UAV 1 (12:21:26) => Arming for T-13L
UAV 1 (12:21:31) => Arming completed for target T-13L
UAV 3 (12:21:31) => Under Fire from threat H-2
UAV 1 (12:21:36) => Available to fire for target T-13L
UAV 1 (12:21:39) => Firing on T-13L
UAV 3 (12:21:42) => Under Fire from threat H-2
UAV 1 (12:21:45) => Firing completed for target T-13L
UAV 3 (12:21:52) => Under Fire from threat H-2
UAV 3 (12:22:02) => Under Fire from threat H-2

Figure 3-10: The UAV Health and Status Updates Window

3.3.3   UAV Health and Status Updates

The bottom right of the decision support display contains a notification window similar in appearance to the chat box (Figure 3-10). The health and status updates window separates human from system communications, and therefore only contains messages from individual UAVs. This reduces clutter and confusion about the source of incoming messages in the chat box window while still allowing the system to notify the operator of important changes.

The system generates three types of messages in MAUVE: 1) emergency health messages in red, which are accompanied by an auditory alert, 2) important status messages in bold black, and 3) lower priority informational status messages in standard black. The emergency health messages are generated when a UAV is under fire from a threat area. Higher priority status messages are automatically sent by UAVs when they are ready to arm and fire, and the UAVs also send lower priority messages informing the operator of their status during all stages of arming and firing.

59

3.3.4 Decision Support

The decision support, represented by timelines, always appears in the top right of the decision support display (Figure 3-7). The manipulation of the appearance and functionality of this window is the primary independent variable of the experiment that will be discussed in Chapter 4. There are four possible forms of decision support in MAUVE that represent levels 1, 2, 4, and 6 on Sheridan and Verplank's levels of automation scale (Table 2.1), termed manual, passive, active, and super active respectively. This section will discuss the functionality and rationale behind the design of each of these levels, but leave the experimental design details for Section 4.5.

The basic premise of the decision support is to simplify standard air tasking order (ATO) data and combine it in a single interface with up-to-date mission planning information. An ATO provides a schedule of events and required resources needed over a period of hours and/or days for an offensive attack. Examples of information contained in an ATO are which aircraft/UAVs are assigned to certain strikes, times on targets, waypoints that must be flown on those strikes, and call signs to be used on those missions. As air tasking orders often involve a large number of aircraft with multiple missions, they are complex and often hard to interpret, particularly under time pressure. Despite this, operators are still expected to extract the information they need in a timely manner. Figure 3-11 shows part of a typical paper-based ATO from Operation Desert Storm that illustrates how difficult it is to extract relevant information from these tasking documents. Once operators have obtained this information and the mission commences, they must constantly compare the current and projected mission status to the ATO and make adjustments to ensure compliance with mission requirements. Combining these sources of information, which often include external sources as well, is a cognitively demanding task that can easily overwhelm operators under time pressure, particularly when supervising multiple vehicles at once. It is for these reasons that the decision support in MAUVE is directed at simplifying this process. While it is known that some level of decision support is required to more

```
OPER/DESERT STORM//
MSSID/MISREF/CTG 150.11/059/FEB//
REF/A/RMG/USCENTAF HQ/1413562FEB91//
AMPN/AIR TASKING ORDER//
NARR/KTO KILLBOX AG6, CHEMICAL WEAPONS AND TRUCKS//
HEADING/AIR OPERATIONS//
MSNID/MSNNO'S 0341I, 0343I, 0345U, 0351P, AND 0353V//
EVTASK/CV61/152315Z/160158Z/CV61//
HEADING/TARGET STRUCK-SIGHTED//
1MSWTGT
/DE/TGT-ID      /LOCATION        /DMPI
/01/            -/294248N0473786E/MILITARY TRUCKS
/02/            -/294300N0473806E/TRUCKS WITH POSS WEAPONS
/03/            -/294300N0473885E/PROBABLE TRUCKS
/04/            -/294236N0473786E/2 MILITARY TRUCKS//
AMPN/THE STRIKE PACKAGE FLEXED TO THIS TARGET IN KTO KILLBOX AG6
PRIOR TO LAUNCH. THE TARGET AREA CONSISTED OF PROBABLE CHEMICAL
WEAPONS AND TRUCKS.//
1TGTRAM

MID      INFO    OPS                                           ROUTE V2173
         CO      XO      STROPS   INTEL    CDC       NAV        SURFAC
         EWO     COMMO   ASW      WPC
CCGS     INFO    N2-40COP                                      ROUTE V09/A
                                                                21-7K1
         01-002  01-003  01-004   01-31    01-32    01-33      01-30
         01-55   01-36   M1-TOTAL MID CO   MIDWPC   MIDCOM

PRN 15295 COPIES 019/034          GIAA 00186         CTG ONE FIVE FLUX AT
TOR 160430Z TOP 160430Z           RUNGSGU 0551  (/2)        160350Z FEB 91

PAGE  1 OF  2
                                  UNCLASSIFIED
```

DECLASSIFIED
ON: 18 OCT 96
BY: DON GWDP IAW
OPNAVINST 5513. SERIES

Figure 3-11: Example High-Level ATO Document

effectively manage ATO information and scheduling, it is not clear what level of automation will provide the most improvement in overall schedule maintenance while avoiding negative side-effects, such as a loss of situation awareness. Therefore, four versions of the decision support were created and structured so that higher levels of decision support expanded upon the features found in lower levels while still retaining all of the functionality and basic information content from previous levels.

The fusion of basic ATO information with current mission planning information resulted in a decision support structure with both hard and soft constraints. As a mission plan is continually modified by operators in response to events unfolding in the battlespace, it does not necessarily have to satisfy the ATO, so the decision support was designed to compare the ATO to the current mission plan. Hard constraints built into the decision support come from the ATO and include such information as the set of targets for the current mission, the TOT windows or deadlines for each of those targets and whether or not the target requires BDA. ATO elements can change occasionally during discrete re-planning events, but as far as the operators are concerned, these are static elements not under their control. Soft constraints on the mission can be directly manipulated by operators and includ elements such as

the set of targets each UAV is due to visit, estimated time of arrival (ETA) at all navigation points, and whether BDA would be performed at each individual target. This information is shown separately for each UAV.

Manual

The manual level of decision support (Figure 3-12, LOA 1) presents all required ATO and mission planning information in a text-based table format. Under the "Current Target" and "Upcoming Active Targets" headings, current TOT windows and ETAs for up to the next 4 targets in queue are presented next to each other for easy comparison. ETAs for arrival back at base in the "Mission Finish" column are listed as well as the next waypoint or navigation point on the current route segment (if applicable) under "Next Waypoint or Loiterpoint". Further assistance is provided to the operator through the "Next Expected Action" column, which tells the user what they should be doing next and at what time, according to the ATO and mission plan depicted on the map display. All of this information is updated dynamically to reflect changing ATO requirements and mission planning changes initiated by the operator. With some effort, all of this information could be mentally constructed to create an accurate future schedule for each UAV. For example, in Figure 3-12 the decision support for UAV 3 tells the following story: UAV 3 will arrive at waypoint WP-3A at 12:05:48, before moving onto target T-12L and arriving at 12:07:41, almost 1.5 minutes before the start of that target's firing window at 12:09:15. The next two targets in the queue are T-4H and T-10H, and the UAV's current ETAs at those targets are both before the start of the applicable firing windows, so this UAV currently is ahead of schedule. Assuming everything goes according to plan, the UAV will arrive back at base at 12:26:10. This level of decision support is termed manual because the automation does not take any decisions or actions for the human. The ATO is simply organized in a digital format for easer reading as compared to a paper version (Figure 3-11).

UAV	Current Target TOT / ETA	Next Waypoint or Loiterpoint	Next Expected Action	Upcoming Active Targets TOT / ETA			Mission Finish
1	T-6M 12:05:55Z - 12:06:15Z 12:05:34Z	-	Arming for T-6M 12:05:45 - 12:05:55	T-3H 12:13:55Z - 12:14:15Z 12:13:22Z	T-13L 12:18:10Z - 12:18:30Z 12:17:27Z	-	Base 1 12:23:36Z
2	T-16H 12:06:35Z - 12:06:55Z 12:05:55Z	-	Arming for T-16H 12:06:25 - 12:06:55	T-14L 12:10:20Z - 12:10:40Z 12:09:49Z	T-11M 12:15:25Z - 12:15:45Z 12:14:56Z	T-8L 12:20:25Z - 12:20:45Z 12:19:56Z	Base 1 12:23:00Z
3	T-12L 12:09:15Z - 12:09:35Z 12:07:41Z	WP-3A 12:05:48Z	Arming for T-12L 12:09:05 - 12:09:35	T-4H 12:11:05Z - 12:11:25Z 12:10:28Z	T-10H 12:18:55Z - 12:19:15Z 12:18:11Z	-	Base 1 12:26:19Z
4	T-17M 12:08:05Z - 12:08:25Z 12:07:53Z	WP-4A 12:06:46Z	Arming for T-17M 12:07:55 - 12:08:25	T-1L 12:10:55Z - 12:11:15Z 12:11:04Z	T-5H 12:16:10Z - 12:16:30Z 12:15:42Z	T-9L 12:20:35Z - 12:20:55Z 12:20:06Z	Base 1 12:25:19Z

Figure 3-12: Example Manual Decision Support Window

Passive

The passive LOA (Figure 3-13, LOA 2) takes all of the ATO and current mission information contained in the manual level and transforms it into a horizontal timeline format color coded by action (Table 3.1). The major difference between this and the manual level is that the entire schedule is visually integrated with both color and iconic representation. These are display design interventions that reduce cognitive load by allowing operators to directly perceive relationships instead of deriving them through calculation, inference, and data transformation. Spatial relationships and actions are understood better with visual representations as opposed to linguistic [37], thus this passive display should improve performance over that of the manual display.

The visual timelines are relative and represent time up to 15 minutes into the future. Figure 3-14 illustrates the standard elements of a representative visual timeline. As can be seen from the diagram, target ETAs are represented by black rectangles on the bottom of each timeline, and waypoint, loiter point and base arrival times are

63
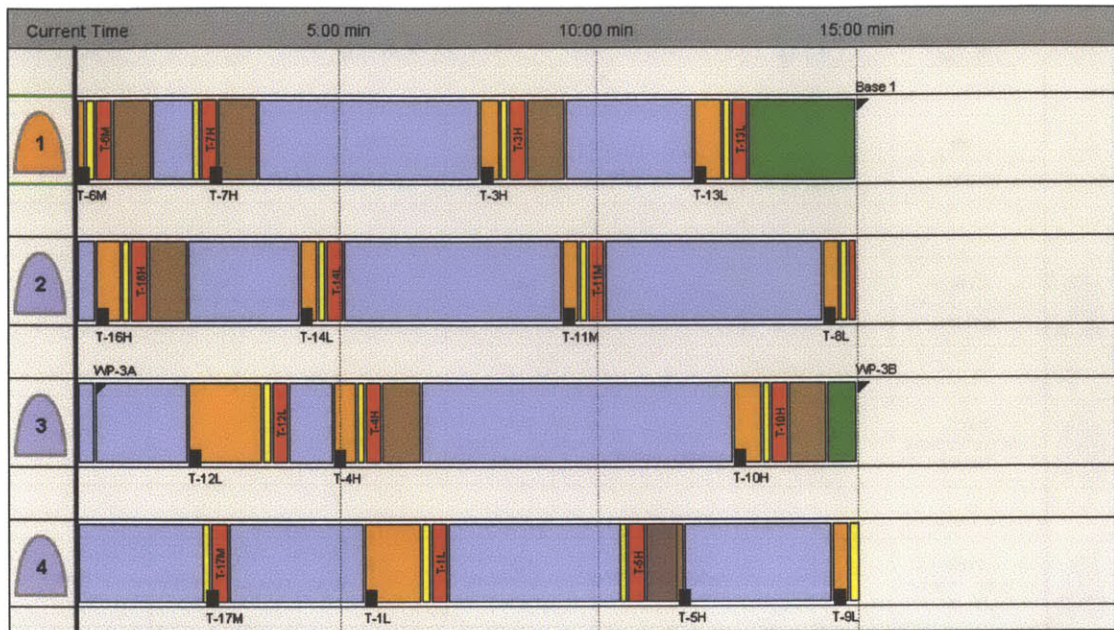
Figure 3-13: Example Passive Decision Support Window

marked by black triangles on the top of each timeline. ATO elements such as target TOT windows, arming windows, and times when BDA is required are represented by red, yellow and brown blocks of time at the appropriate times.

With this visual representation, recognizing problems with the current mission plan becomes a process of comparing the relative location of display elements instead of specific times to one another. The visual task is much easier and faster to perform because it only involves perception, as opposed to the cognition required to compare the ETA and TOT numbers to each other. An example of how users recognized problems with the mission plan using this visual timeline is as follows: in Figure 3-13, the ETA marker for target T-5H (assigned to UAV 4) is entirely to the right of that target's firing window. This means that UAV 4's arrival time to T-5H is after the deadline for that target, and a upon seeing this, users knew right away that they were going to miss that target and that they had to act to change this. Another emergent feature of this display is that if the color of the UAV icon to the immediate left of the timeline is not the same color as the current timeline block, the operator immediately knows that the UAV needs attention, as this means that the UAV is missing a target
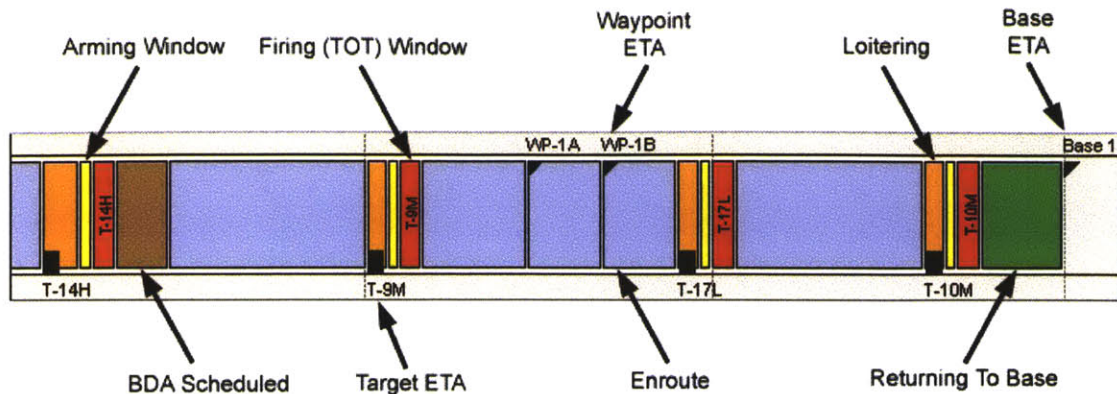
Figure 3-14: Representative Decision Support Visual Timeline

deadline specified by the ATO. This level of decision support is termed passive because the automation is not performing any tasks except transforming the basic ATO and mission planning information into a graphical format.

Active

The active LOA (Figure 3-15, LOA 4) uses the same horizontal timeline format as the passive automation level, but provides additional computer aid to the user. In this version of the decision support, an algorithm searches for periods of time in the schedule that it predicts will cause high workload for the operator, and it directs the operator's attention towards them. The computer identifies high workload areas, or "bottlenecks" as periods of time during which multiple UAVs are scheduled to be simultaneously executing mission critical actions, defined as arming, firing, or performing BDA. The automation draws attention to these areas of concern by a reverse shading technique, in which the "bottlenecks" are highlighted while the rest of the timeline's colors are muted into the background, but still visible. As no information is hidden, only made less salient, the operator's attention is directed to the appropriate areas of the schedule while allowing them to maintain SA for the rest of the mission. This technique also does not require any drilldown to subsequent
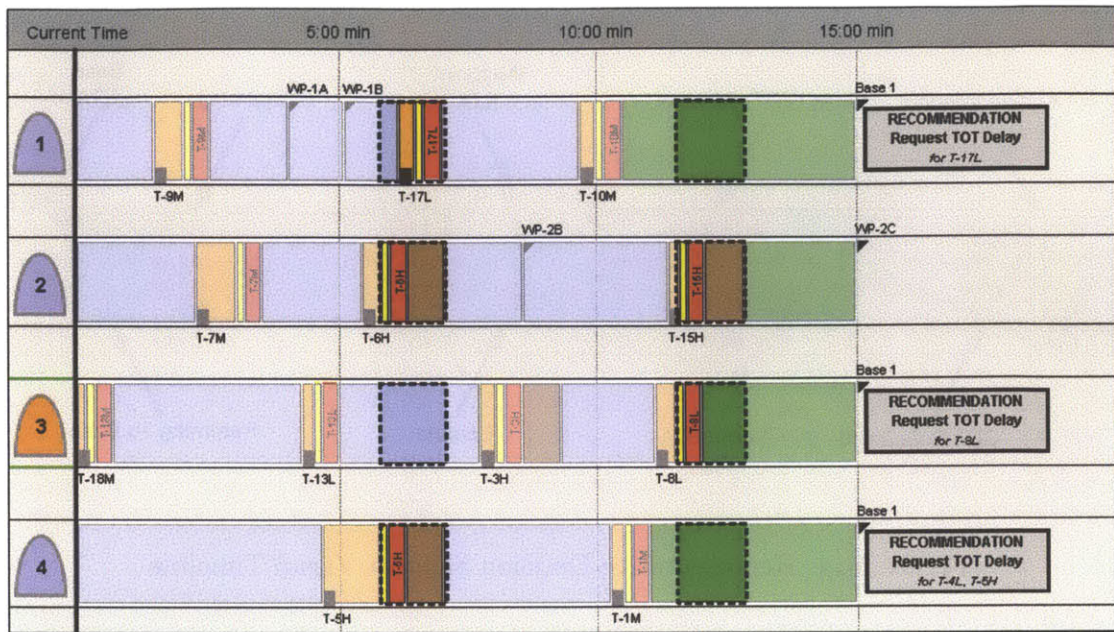
Figure 3-15: Example Active Decision Support Window

screens to solve the problem, so it permits operators to make local changes to alleviate workload and immediately see their effect on the global mission plans of all UAVs.

In addition to identifying areas of high workload, the computer also recommends a course of action to spread out the operator's workload, such as the delay of a particular TOT. Computer recommendations appear in gray boxes to the right of each UAV's timeline that they pertain to (Figure 3-16). It should be noted that while the automation makes locally optimal recommendations, the algorithm is not globally optimal. In other words, following the computer's recommendation to relieve a high workload area always removes that particular schedule conflict, but sometimes it creates another in the process.

Schedule changes to relieve areas of high workload are requested by the operator if they so desire (Section 3.2.3). The active level of decision support gives subjects several options: 1) they could acknowledge a high workload area but take no action, 2) they could follow the recommendation to relieve the projected high workload area by shifting a TOT, or 3) they could make other mission planning changes to ensure the high workload area does not occur in the first place, such as removing a target

66

RECOMMENDATION
Request TOT Delay
for T-17L

Figure 3-16: Automation Recommendation for a TOT Delay Request

from a UAV's mission plan. If all high workload areas are taken care of by the operator, full color is restored to the timeline and the automation looks and acts like the passive level of automation until such time that a schedule change creates a new high workload area.

An approved TOT delay request causes the deadline for the relevant target to automatically be moved back in one of the following ways: 1) 30 seconds, or 2) far enough back to prevent that target's arming, firing or BDA windows from occurring in the same time period as any other overlapping targets. For example, if a target has a firing window that overlaps with two other high priority targets assigned to other UAVs, an approved TOT request for that target would move it back 45 seconds or more to avoid the start of that target's arming window overlapping with the other targets' BDA windows. In this way, TOT requests occasionally move a target's TOT window back a minute or more. The algorithm also ensures that deadline moves do not create conflicts with other targets assigned to the same UAV.

The purpose of this level of automation is to help operators to identify time periods of potential concern where they could be overwhelmed farther in advance, so that they can better plan to avoid them, or at least be better prepared to handle them. This level of decision support is termed active because the automation narrows down a set of possible solution alternatives for high workload problems to a single recommendation, consistent with Sheridan's level 4 of automation.

Super Active

The super active LOA (Figure 3-17, LOA 6) also builds upon the passive level visual timeline, but instead of making recommendations to the operator as in the active LOA, a management-by-exception approach is taken whereby the computer automat-
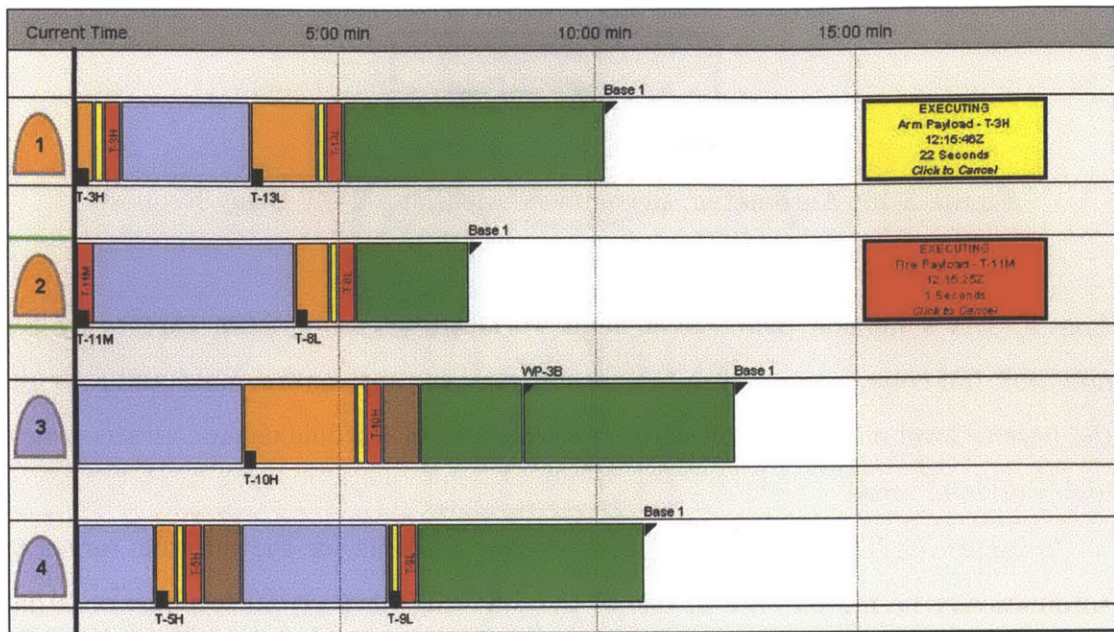
Figure 3-17: Example Super Active Decision Support Window

ically executes the arming and firing actions for all UAVs at each target, when the rules of engagement for such actions are met (Appendix A). For example, in order to fire, a UAV has to be located at the particular target it is due to fire on, already armed, and the scenario time within the TOT window for that target. It should be noted that there are no special rules of engagement for the super active level of automation; the computer follows the same rules for humans in other LOAs. In this way, responsibility for executing routine arming and firing actions is removed from the operator, rendering the computer recommendations of the active level unnecessary.

However, the operator is still responsible for determining if the arming and firing actions are appropriate, and for re-planning actions and manipulating routes to ensure the UAVs arrive at the correct targets on time. Up to 30 seconds in advance before every arming and firing action, exception boxes (Figure 3-18) appear to the right of the timeline that allow the operator to veto these actions. These exception boxes show which target the UAV is due to fire on, which action the UAV is preparing to take, when it will take it, and it contains a countdown timer that indicates how many seconds from current time that the action will be executed. The color of the

EXECUTING
Arm Payload - T-3H
12:15:46Z
22 Seconds
Click to Cancel
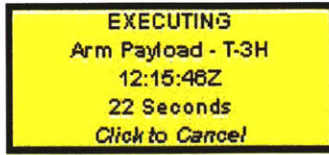
Figure 3-18: An Arming Exception Box Under Super Active Automation

box redundantly indicates which action the UAV is preparing to perform: red for firing and yellow for arming. This color also matches the next timeline block in the relevant timeline. This level of decision support is termed super active because the automation is performing all of the mission critical execution actions for the user.

Summary

Figure 3-19 shows the progression of the decision support window as the level of automation selected in the MAUVE program moves from manual to super active.

(a) Manual

(b) Passive

(c) Active

(d) Super Active
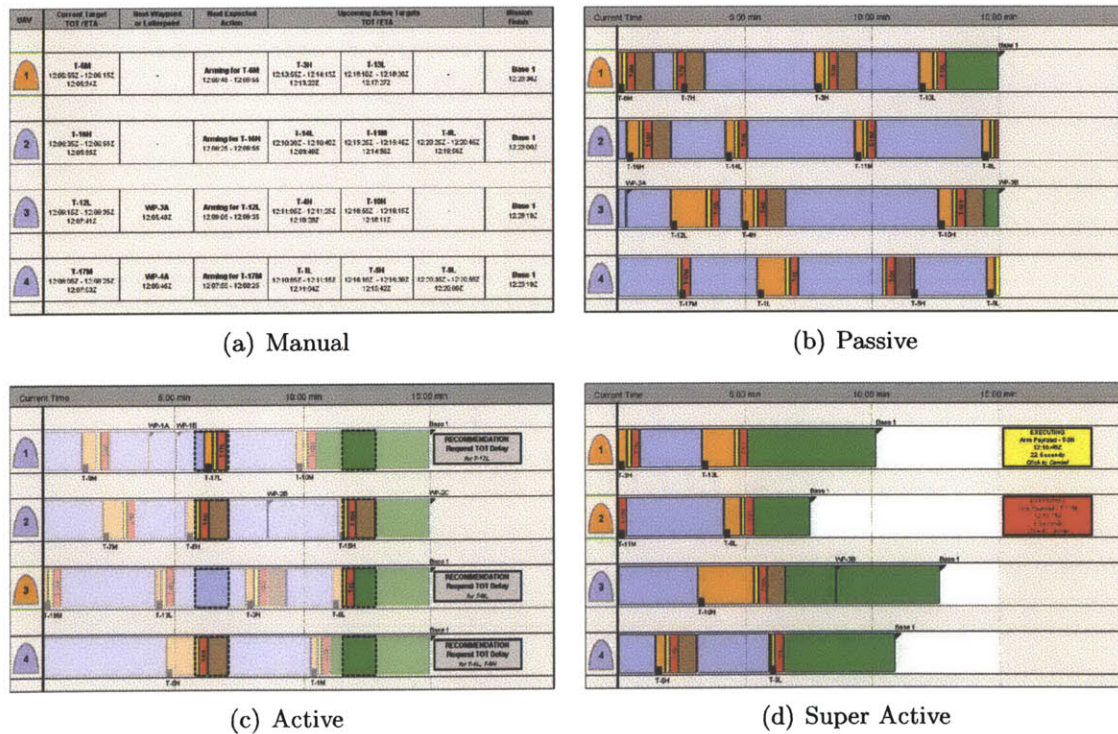
Figure 3-19: The 4 Possible Levels of Decision Support in MAUVE

69

Chapter 4

Hypotheses and Methods

The first part of this chapter presents hypotheses where predictions are made for trends in system performance for each of the four levels of decision support in the MAUVE simulation program and interface described in Chapter 3. The last half of the chapter provides the details of an experiment conducted with MAUVE evaluating different levels of automation in the context of both human and system performance. Major areas of the experiment that are outlined in this chapter are the experimental objectives, apparatus, participants, and the experimental design.

4.1 System Performance Hypotheses

The discussion regarding the effects of automation on wait times in Chapter 2 can now be extended to make predictions about the specific levels of decision support in the MAUVE interface. This section describes the features of the MAUVE simulation and interface that have an impact on wait times, and makes wait time predictions based on these features and the theory introduced in Chapter 2.

4.1.1 Interaction Wait Time (WTI)

Interaction time in MAUVE is likely to be significantly influenced by the timeline introduced for higher levels of decision support, as it provides a visual, color-coded

71

representation of critical temporal information needed to properly task the UAVs. Vertical stacking of the timelines and status display make it possible to observe all vehicle actions simultaneously. As the operator re-plans, the timeline automatically updates thus allowing problem spaces to be more easily explored. All of this information is available under manual automation, but it cannot be obtained without substantial cross-checking of table entries, and the table format does not facilitate schedule comparisons across different vehicles. Therefore, due to the increased visibility and ease of determining temporal relationships between different ATO elements on the visual timeline, WTI should be lower for the passive, active and super active levels of automation in comparison to the manual level.

As the active level of automation further guides vehicle selection and planning (components of WTI, see Section 2.3.2) by drawing attention to predicted high workload blocks of time and making recommendations for specific UAVs to alleviate workload bottlenecks, the total WTI for active automation should be less than the passive level. Under super active automation, planning times for expected events such as arming and firing will be decreased to near zero, as the UAVs autonomously perform all major tasks but re-planning for the human. As a consequence of the high level of automation, the vehicle selection decision is also taken away from the human, thus decreasing subtask selection to nearly zero as well. Therefore, in combination with the benefits of the visual timeline, super active should have the lowest total WTI of any of the levels of automation. Combining all of these predictions together, this means that there should be a decreasing trend of WTI with increasing level of automation. Total WTI should increase with higher levels of re-planning, as a higher number of events occurring in a scenario of similar difficulty means that there will be more opportunities for WTI to accumulate.

4.1.2 Wait Time in the Queue (WTQ)

As outlined in chapter 2, WTQ depends heavily on WTI, which is expected to decrease with increasing levels of automation in MAUVE. Therefore, WTQ should follow the same decreasing trend as level of automation increases. WTQ should also increase

with greater levels of workload and uncertainty because again, WTQ follows the same trends as WTI, but also because at higher levels of re-planning, events occur at a faster pace. Therefore, there should be a greater number of opportunities for queues of tasks to occur.

4.1.3 Situation Awareness Wait Time (WTSA)

As stated in Section 2.3.2, the super active level of automation should eliminate any wait time due to the loss of SA for planned events. This is because under this level the system autonomously carries out the mission plan, so long as the mission plan guides the UAVs to the correct targets at the right times. When unexpected events occur, operators may have low levels of SA due to boredom or complacency, and therefore they may incur WTSA by not noticing that the mission plan needs adjustment or that a UAV has stopped progressing along the mission plan and needs attention. In contrast, under the manual level of automation in MAUVE, the operator is responsible for both execution and re-planning. With very little automation assistance, these dual responsibilities increase workload, which could induce low levels of operator SA. As SA decreases, there is no guarantee that wait times will occur, only that the probability of WTSA occurrence is more likely. However, given the greater number of opportunities for this human error to occur under manual decision support, it should have the largest total amount situation awareness wait times amongst all LOAs, while the super active level should have comparable WTSA. The active and passive levels should have the lowest accumulated WTSAs, as the moderate levels of workload (as compared to manual and super active) provided by these LOAs should cause operators to have an optimum level of SA. WTSA should increase with higher levels of re-planning, as the additional workload placed on operators due to these events will induce lower SA.

4.1.4 Summary

In general, increasing levels of decision support in MAUVE should result in lower aggregate levels of WTI and WTQ, while WTSA will be lower in the passive and active modes (Figure fig:WTTrends). It is not clear which of these components dominates total wait time, or even if a single component will dominate across all factor levels. Increasing workload and uncertainty introduced through higher levels of re-planning will likely increase WTI and WTQ, while WTSA may be contextually dependent.
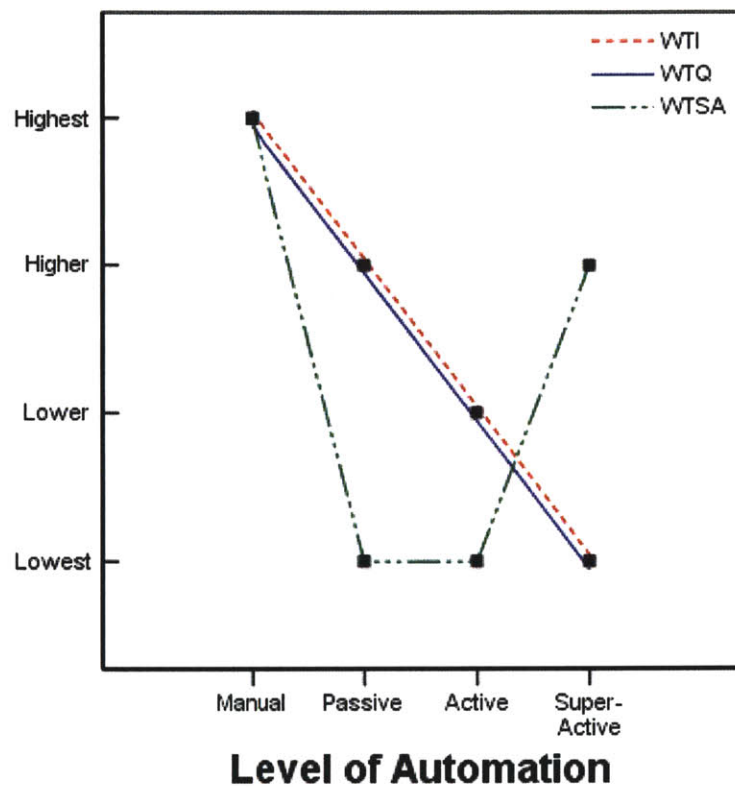

Figure 4-1: Predicted Trends for Wait Time Components Across MAUVE LOAs

4.2 Experiment Overview

In order to address the hypotheses from Section 4.1, an experiment with the MAUVE simulation interface (Chapter 3) was conducted. The goal of the experiment was to

explore how human decision making in a time-critical, HSC multiple task management context affected both human and system performance, and how different levels of automation interacted with these measures. Through this experiment, the wait time predictions as a function of level of automation in Chapter 2 could be validated and the impact of system performance on human performance and vice versa quantified.

4.3 Apparatus

Training and testing was conducted on a four screen system called the multi-modal workstation (MMWS) [28], originally designed by the Space and Naval Warfare (SPAWAR) Systems Center as a test prototype to aid the development of human-computer interface recommendations for future Navy C^2 systems (Figure 4-2). The top three screens used were 21" and were run at 1280 x 1024 pixels, 16-bit color resolution, while the bottom screen was 15" and was run at 1024 x 768 pixels, 32-bit color resolution. The workstation was a Dell Optiplex GX280 with a Pentium 4 processor and an Appian Jeronimo Pro 4-Port graphics card. Subjects interacted with the simulation via a Logitech MX500 cordless mouse and a generic numeric key pad. This is a typical set-up for the MMWS as described by Osga et al [28]. During testing, all mouse clicks and both message box histories, including incoming and outgoing messages, were recorded by software. In addition, screenshots of both simulation screens were taken approximately every two minutes, all four UAV locations were recorded every 10 seconds, and a whenever a UAV's status changed, the time and change made were noted in the data file.

The center top and top right screens contained the simulation navigation and decision support windows, respectively (Figures 3-2 and 3-7). The top left screen contained subjects' objectives in rank priority order for the scenarios (Section 4.5.3, also seen in Appendix B), and was static throughout the experiment, while the bottom screen contained the color coding for UAV actions in the simulation as a reference (Table 3.1, also seen in Appendix B). This information was provided during testing because pre-test feedback indicated it was a useful reminder, and to ensure that

Figure 4-2: The Experimental Set-Up on the MMWS

participants were not required to memorize large amounts of information.

4.4 Participants

A total of 12 subjects took part in this experiment, 10 men and 2 women. Subjects were recruited based on whether they had UAV, military and/or pilot experience. The subject population consisted of a combination of students, both undergraduates and graduates, as well as those from the local reserve officer training corps (ROTC) and active duty military personnel. All were paid $10 an hour for their participation. In addition, a $50 incentive prize was offered for the best performer in the experiment.

The age range of participants was 20 - 42 years with an average age of 26.3 years. 9 participants were members of the ROTC or active duty USAF officers, and included seven 2nd Lieutenants, a Major and a Lieutenant Colonel. Correspondingly, the number of years of military service ranged from 1-20 years, with 3 being the median.

While none had large-scale UAV experience, 9 participants had piloting experience. The average number of flight hours among this group was 120. See Appendix C for a summary of the study demographics.

4.5 Experimental Design

4.5.1 Independent Variables

Two independent variables were of interest in this experiment: level of decision support, and level of re-planning. The level of decision support was a between-subjects variable and the level of re-planning was a within-subjects repeated variable, so subjects were randomly assigned to a LOA factor level but experienced all re-planning conditions.

Level of Decision Support

As discussed previously in Chapter 3, to assess the impact of increasing levels of automation on human and system performance, various forms of decision support were provided to subjects that represented levels 1, 2, 4, and 6 on Sheridan and Verplank's levels of automation scale (Table 2.1), termed manual, passive, active, and super active respectively. The main differentiating features between the different levels were:

- **Manual** - The ATO and mission planning information was presented in an organized, digital table format for each UAV.

- **Passive** - The ATO and mission planning information was transformed into a visual timeline format for each UAV.

- **Active** - The passive visual timeline was used to convey ATO and mission planning information. Additionally, the computer identified areas of potential high workload on the timeline for subjects and made recommendations to relieve them.

- **Super Active** - The passive visual timeline was used to convey ATO and mission planning information. The computer took over automatic execution of all arming and firing actions, and provided operators with the ability to stop an automated action within 30 seconds of notification.

This was the primary independent variable for this experiment. For more discussion and details on the various interface layouts and functionality, see Section 3.3.4.

Level of Re-planning

The secondary independent variable was the level of re-planning, either low or high, which represents operational tempo. It has been demonstrated that in supervisory control of multiple autonomous vehicles, increased operational tempo can be a significant contributor to degraded human performance [7]. Schedule re-planning was required if any of the following events occurred:

- An emergent target appeared in the battlespace

- A target was assigned to a different UAV strike mission

- An emergent threat area appeared in the battlespace

- A threat area became inactive

- A battle damage assessment requirement was added to an existing target

- A battle damage assessment requirement was removed from an existing target

- A system failure occurred, requiring a UAV to return to base unexpectedly

Low and high levels of schedule re-planning were investigated. The low re-planning condition contained 7 re-planning events, while the high re-planning condition contained 13. Groups of re-planning events were interspersed at approximately 3 minute intervals, but under the low re-planning level these groups only ever consisted of a single event. Under high re-planning, some groups were composed of 2 or 3 re-plans occurring within 60 seconds of each other.

4.5.2 Dependent Variables

Several dependent variables were used in this experiment: a performance score, wait times as discussed in section 2.3.1, situation awareness, critical events, and secondary/subjective workload measures.

Performance Score

In order to adequately measure an overall level of performance for a given test session, a performance score was created that incorporated how well subjects met the numerous objectives for that session. The performance score was a product of the targets correctly destroyed, including their priority and difficulty level, and number of times BDA was correctly performed. The ATO specified that some targets in the scenarios should not be destroyed, and so operators were significantly penalized for erroneously firing on these targets. Penalties were also assessed for hits taken by UAVs in threat areas and for having UAVs return to base beyond the mission time limit. A score of zero on this rating scale indicated no objectives had been met in the session, while a perfect score of 1000 meant that all mission objectives were met. For complete details on the scoring formula and rationale behind its development, see Appendix D.

Wait Times

As outlined in Section 2.3.1, overall system wait time is composed of interaction wait time (WTI), wait time in the queue (WTQ), and situation awareness wait time (WTSA). In the context of the MAUVE simulation, system wait times were accumulated whenever one or more vehicles were in a degraded state such that their activities were counter to or not advancing the mission plan. This degraded performance could have taken the following forms in the simulation:

- Unnecessary loiter time at a target, such as if a UAV was loitering at an active target, but failed to fire before the end of the firing window, or if a target was removed from the mission plan while the UAV was loitering at that target, and any time spent loitering at a target while available to fire without actually doing

so. In all cases, the UAV continued to loiter at the target indefinitely until the operator issued it a subsequent command.

- Any time from the beginning of any arming window until the UAV began the actual process of arming. For missed targets, a full 30 seconds (10 second arming window plus 20 second firing window) of wait time was assessed, as during this time, people on the ground expected a weapon on target and it did not happen. In this case, the wait times incurred can be thought of as times the broader system was affected by the UAVs degraded performance. Additionally, if a UAV missed performing BDA on a target it was supposed to, a 45 second wait time was added to that UAV's total.

- Any time the UAV was performing incorrect actions, such as arming and firing on a target removed from the mission plan by the ATO or performing BDA on unnecessary targets. This also included time when UAVs were flying without a specific purpose or goal in mind, such as when subjects were confused about which target to fly to next and had the UAV moving in circles, back on itself, etc.

- Time spent in threat areas, until the operator paid attention to the UAV and planned a (reasonably) optimal path out of the threat area.

- The time period from when a return to base command was issued for a UAV until the order was followed. This was a wait time because a RTB command represented a system failure and the UAV could no longer perform its required tasks for the mission.

From the recording of control activations, UAV locations, screenshots, and observer notes taken during trials, an accurate "story" of the test sessions could be re-constructed in later data analysis. From this, each of the wait times incurred was classified into WTI, WTQ or WTSA. In general, a wait time was classified as WTI if the UAV was receiving operator attention, WTQ if the operator was busy with

another UAV and transitioned to the relevant UAV quickly after finishing that task and WTSA otherwise.

Situation Awareness

As discussed in Chapter 2, situation awareness (SA) is generally defined as the perception of the elements in the environment within a volume of time and space, and the comprehension of their meaning and the projection of their status in the near future. SA has three distinct levels, which are: 1) the perception of the elements in the environment, 2) the comprehension of the current situation, and 3) the projection of future status [13, 14].

Situation awareness in MAUVE was captured both objectively and subjectively. The objective measure was obtained by how accurately subjects answered 12 online questions asked at approximately 2 minute intervals in the chat box window. In order to avoid confounding the responses with typing ability, all questions were designed to have very short numerical answers a maximum of 4 keystrokes in length. Examples of the types of information that were required by these communications are as follows:

- Which UAVs will arrive at their next navigation point by 12:05:00?

- If UAV 1 was issued a RTB command, which of its currently assigned active targets could be most easily re-assigned and destroyed by UAV 4?

- How many targets have TOTs starting in the next 5 minutes?

- According to the current mission plan, how many UAVs will arrive at Base within 2 minutes of UAV 4's arrival?

- Which future high priority target would provide the greatest decrease in a single UAV's travel time, if removed from the ATO?

Six questions per session were designed to assess level 2 SA and the remaining six level 3 SA. A question was considered a level 2 SA type of question if the answer could be directly perceived from the decision support or map display, either right away or

with a minor amount of data synthesis, while level 3 SA types of questions required greater amounts of data synthesis and projection of current system states to future states.

The answers to each question were graded on a 0-5 scale for correctness, with zero points received if no response was received and 5 if the correct answer was given. Intermediate scores of 1-4 were assigned based on the degree of correctness of the answer given, and each question was also weighted on a scale of 1-5 to take into account difficulty. The question weighting was based upon how many steps were required to solve the problem, mediated by the context in which the question was asked. Questions asked during periods of high workload were more difficult, and sometimes question difficulty was influenced by scenario events. For example, the question "How many targets have TOTs starting in the next 5 minutes?" became harder to answer if scenario events were manipulated by a subject so that a target's TOT deadline expired soon after the question was asked. Typically this was an easy question because it could be directly perceived from the decision support, but the decision support did not show deadlines in the past. In this case, subjects coming back to answer the question at a later time had to read the decision support as well as remember they destroyed or missed a target in the time since the question was asked. Each response to a question was multiplied by the weighting for that question and an aggregate score obtained before being divided by the total possible score. In this way, an average SA score on a scale of 0-5 was obtained. The questions asked were equally divided between those testing level 2 SA (data comprehension/synthesis) and level 3 SA (projection) and SA scores for these categories were separately examined.

A subjective SA scale constructed from expert observer ratings was also constructed to test the validity of the objective SA measure, as subjective SA scales based upon expert observer ratings have been found to be a reliable and valid measure way to measure SA [2]. This scale was based upon expert opinion on what were judged to be the four most salient indicators of SA in the test sessions:

1. The number of entries into threat areas where the UAV received 3 or more hits and the operator did not intervene to minimize further damage to that UAV.

2. The amount of system wait time at targets due to loss of situation awareness.

3. The number of targets missed due to lack of situation awareness.

4. The percentage of re-planning events successfully completed.

The first two indicators measured level 2 SA (comprehension), as a UAV loitering unnecessarily at a target or a UAV traveling through a threat area could be perceived directly from the map display. The second two indicators measured level 3 SA (future projection), as a failure to successfully complete a re-plan or a missing a target due to a lack of SA meant that the operator had not properly executed, modified and projected the mission plan in response to scenario events. Each test session was assigned four individual scores, one for each indicator, and the average of these ratings was the subjective SA score for that test session. A score for Level 2 SA was obtained by averaging the ratings for indicators 1 and 2, while a score for Level 3 SA was obtained similarly with indicators 3 and 4. For more details on the development of the subjective SA scale used in this study, see Appendix E.

Critical Events

An error of commission occurs because a wrong action is taken, in contrast to an error of omission which occurs as a result of not performing an action. Operators could commit errors of commission in the test sessions by erroneously firing upon targets that the ATO had previously specified should not be destroyed. These were considered important and critical events because errors of commission in military applications often have potentially serious consequences, such as firing on civilians or friendly forces. As an example, the US Army's Patriot missile system, often run in highly automated modes corresponding to LOA 6 (Table 2.1), is particularly susceptible to operators committing errors of commission. In the most recent accident with this system, during the 2004 war with Iraq, a Patriot missile system engaged in fratricide, shooting down a British Tornado and an American F/A-18, killing three. The number of times that critical firing events occurred during test sessions was recorded for subsequent data analysis.

83

Workload Measures

The use of response times to the online chat questions as a measure of secondary workload has been shown to be an effective technique [7], and was incorporated into this research. To measure subjective workload, after the completion of each test session, subjects filled out a modified NASA Task Load Index (TLX) subjective workload rating survey. The NASA TLX rating scale has been tested in numerous experimental conditions and has been found to be a reliable indicator of subjective workload [19]. The standard NASA TLX procedure computes a single workload score from subjects' weighted ratings on a 1-20 scale along 6 dimensions, which are mental demand, physical demand, temporal demand, effort, performance, and frustration. As the experimental task required no physical demand, subjects were told to ignore parts of the survey asking about that dimension and the subsequently modified subjective workload scores were based upon the 5 remaining dimensions.

4.5.3 Test Scenario Design

Objectives

Subjects had two main objectives in this experiment: 1) To guide each UAV's actions so that together, all UAVs under their supervision properly executed the required missions of the ATO, which could change over time, and 2) To answer periodic questions about the situation from commanders through the instant messaging tool. Supervision of the UAVs was broken down into the following prioritized sub-tasks, from highest priority to lowest:

1. Return to base (RTB) within the time limit for the mission.

2. Comply with changing mission requirements, which will be relayed by periodic intelligence messages, such as a RTB order earlier than the time limit.

3. Destroy all targets before their time on target (TOT) window ends.

4. Perform battle damage assessment (BDA) on specified targets after destroying them.

5. Avoid taking damage from enemies by navigating around and out of threat areas.

6. Answer communications.

These sets of objectives often conflicted with one another. In these cases, subjects were required to perform the actions that had the highest priority at the expense of lower priority objectives.

Test Scenarios

Two test scenarios were designed for this experiment. The only significant difference between them was the number of re-planning events that occurred, as described in Section 4.5.1. Base difficulty before taking into account re-planning events was approximately the same, and was controlled by having the same number (19) and priority of targets (6 low, 7 medium, 6 high) to be destroyed by the UAVs in the same amount of time (30 minutes). In both sessions, UAVs initially launched with a pre-planned validated mission plan in which no deadlines would be missed, so long as no unexpected events occurred and the operator executed the required actions at the appropriate times. With this initial mission plan, operators had a specified slack time of 30 seconds before every target, which was kept constant across scenarios and used to control difficulty. Thus, if operators followed the mission plan exactly and no unexpected events occurred, all the UAVs would always arrive 30 seconds before the beginning of the arming window at each target. Initial routings for the UAVs were mirrored and rotated but otherwise kept the same between sessions. Target names were changed and initial target deadlines varied slightly to give subjects the perception that they were seeing different situations. Some changes to routings were necessary to maintain the 30 second slack time rule in order to accommodate different re-planning events across sessions. Figure 4-3 shows how the map display looked to subjects at the beginning of the high and low re-planning scenarios, respectively.
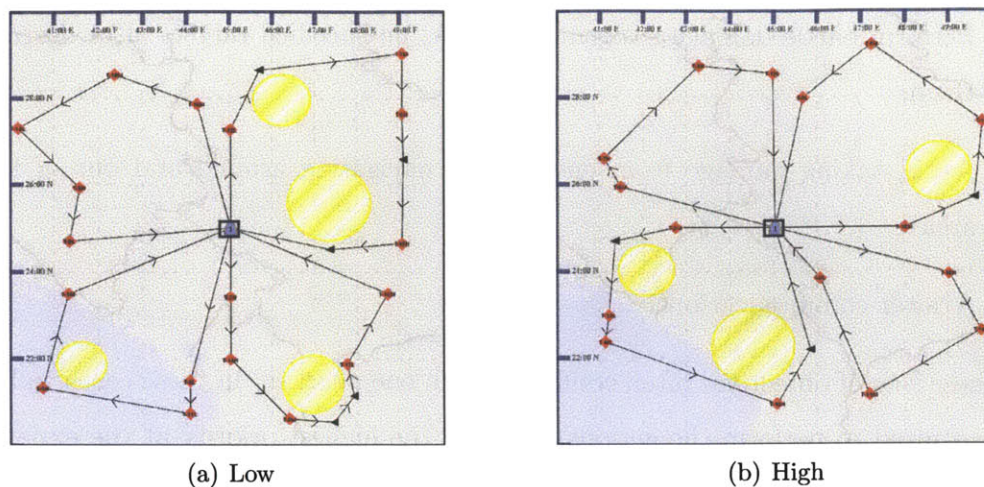
(a) Low   (b) High

Figure 4-3: Test Session Maps

4.5.4 Testing Procedure

Before arriving at their test session, subjects were sent a pre-experiment tutorial in order to familiarize them with the functionality and usage of the MAUVE interface. Individual customized tutorials for each level of automation were created so that subjects were exposed only to the particular version of the decision support they would experience. Upon arrival, subjects filled out a consent form (Appendix F) and a demographic survey (Appendix G) before reading a short instruction sheet that detailed their objectives in the scenarios, example chat box questions, and specific rules of engagement (Appendix H).

All subjects then received between 90 and 120 minutes of training until they achieved a basic level of proficiency in monitoring the UAVs, redirecting them as necessary, executing commands such as firing and arming of payload at appropriate times, and responding to online instant messages. A written testing script was followed and the same experimenter trained all subjects to ensure there was consistency in this training, which was carefully structured into three practice scenarios. The first scenario served to familiarize participants with the basic display elements, rules of engagement, and taught subjects basic mission execution actions. The second practice scenario focused on teaching subjects how to perform mission planning actions,

86

particularly in response to re-planning events. During this scenario, subjects solved every possible type of re-planning event. The third practice scenario was a hands-off 15-minute test during which subjects were able to get a feel for the pacing and timing of the test scenarios. This was also the first session during which subjects were required to answer questions through the chat box while supervising the UAVs. The only differences between the final standard training session and the test sessions were its shorter length and marginally easier difficulty level. After scenario completion, subjects were fully debriefed on their mistakes, and any remaining questions they had were answered. Based upon pre-testing, subjects were required to achieve a minimum pre-determined level of performance to move onto the actual test sessions. In only 2 cases of 12, an additional practice scenario was required, which was similar in content to the previous practice session. Both subjects that received extended training improved enough to meet the established minimum level of proficiency.

Following training, participants tested on two consecutive 30 minute sessions, one with a higher number of re-planning events than the other. The order in which the conditions were presented to subjects was randomized and counter-balanced to prevent a possible order effect. The experimenter observed and took notes on how subjects responded to re-planning events answered questions. Overall behaviors or comments were also recorded, particularly in instances where an unusual situation arose that would have been difficult to interpret from the raw objective data by itself. For an example of typical trial progression, see Appendix I. This Appendix shows both MAUVE displays at approximately 2 minute intervals over the course of an entire test session. At the conclusion of each test session, subjects filled out an electronic NASA TLX survey, and after the both sessions were concluded, feedback on the study experience and interface was solicited.

Chapter 5

Results

This chapter presents statistical findings from the experiment described in Chapter 4, which had 2 independent variables: level of automation and level of re-planning. The dependent variables examined include the performance score, number of TOT delay requests made, proportion of TOT delay requests approved, wait times, situation awareness, critical events, and workload metrics.

5.1 Overview

As this experiment examined two independent factors, 4 levels of automation (between-subjects) and 2 levels of re-planning (within-subjects), the general linear statistical model used for the majority of the analysis in this chapter is a 2x4(3) repeated measures linear mixed model. Three subjects were nested within each automation level, and both independent factors were considered to be fixed while subjects were a random factor. Age was used as a covariate in all analyses, and for all reported results $\alpha = 0.05$ unless stated otherwise. Post-hoc comparisons of were done using equal or unequal variance (depending on the significance of the Levene/Brown Forsythe tests), small population mean t-tests with Bonferroni correction.

89

5.2 Overall Performance Measures

5.2.1 Performance Score

As outlined in section 4.5.2 and Appendix D, the performance score was an aggregate measure of overall human performance that incorporated the number and priority of targets correctly destroyed, number of times BDA was correctly performed, and any penalties from threat hits, late arrivals to base, or targets erroneously destroyed. Figure 5-1 shows the average performance scores for each experimental condition. Level of re-planning was significant ($F(1,9.9) = 19.40$, $p = 0.001$) while level of automation was marginally significant ($F(3,10.8) = 3.04$, $p = 0.076$). There was no significant interaction between the factors.

Further testing revealed that the super active level of automation had significantly higher performance than the active level under the high re-planning condition ($p = 0.032$), while the manual and passive levels did not. However, the manual and passive levels also did not perform significantly worse than the super active level under the same conditions, thus of the four LOAs, active automation caused significantly degraded performance under high re-planning. There was no significant difference between the automation levels under low re-planning.

5.2.2 Time on Target Delays

Based upon observed behaviors in test sessions as well as the statistical results from the performance score analysis, a significant driver of human performance was found to be operators' use of the "TOT Delay" function. Through requesting TOT delays, operators could manipulate target deadlines to spread out their workload and/or make deadlines they otherwise would have missed due to inadequate planning and execution at some earlier time in the scenario (Section 3.2.3).

Counts of TOT delays requested and granted were both examined in non-parametric analyses. A Wilcoxon Signed Rank Test between levels of re-planning found that the number of TOT requests made was significantly lower in the low re-planning condi-
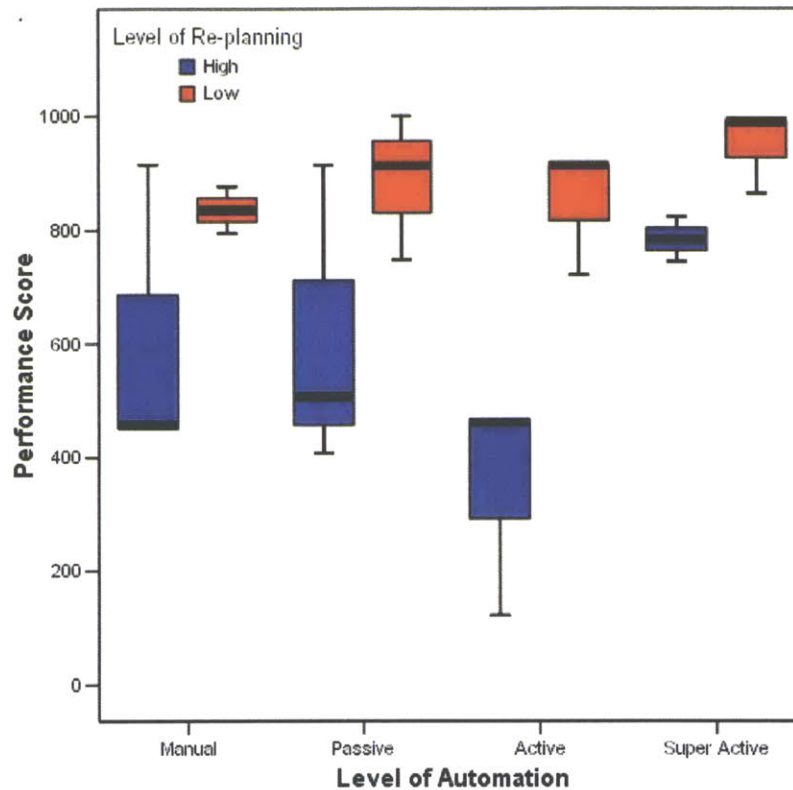
Figure 5-1: Performance Scores Across Levels of Automation and Re-planning

tion (p = 0.059) and not significant for number of requests granted (p = 0.167). A Kruskal-Wallis test showed that the number of TOT delays requested was marginally significant between automation levels (p = 0.096). Inspection of the ranked data and Figure 5-2 reveals that the difference between automation levels in this case was driven by a lower number of TOT delay requests in the manual level of automation (under low levels of re-planning), and a higher number of TOT requests in the active level. Wilcoxon Rank Sum Tests comparing the manual and active levels of automation to the other levels were performed to investigate these hypotheses. This analysis revealed the manual level had significantly lower numbers of TOT requests than the passive level (p = 0.015), with marginal significance in comparison to the active level (p = 0.06). Comparison to the super active level revealed no significant difference in number of TOT requests made. A median chi-square test showed that the number of

TOT requests made under the active level was significantly higher than other levels of automation (p = 0.065).
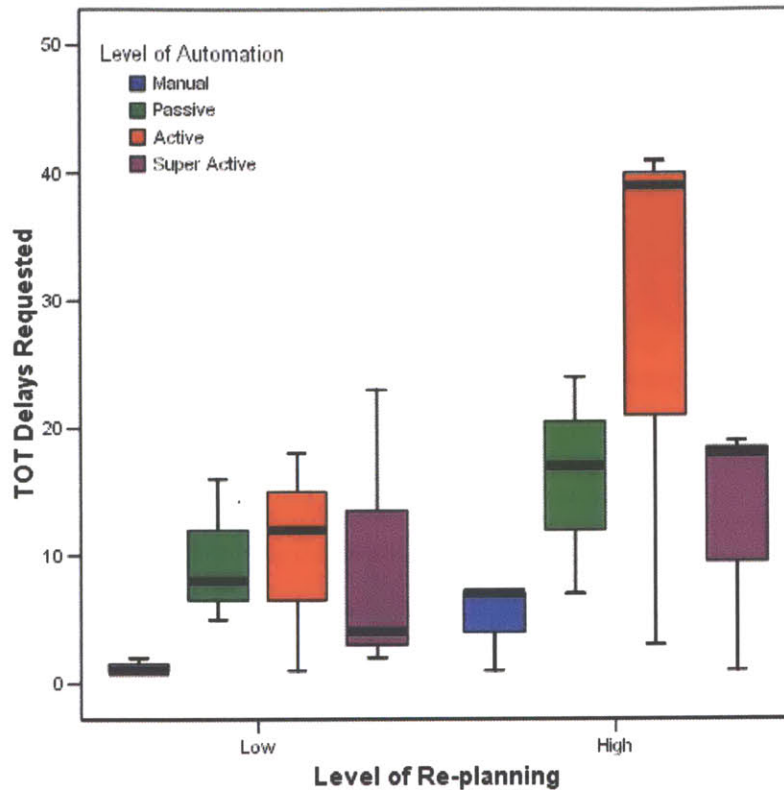
Figure 5-2: TOT Delays Requested Across Levels of Automation and Re-planning

In addition to examining TOT requests, proportion of requested TOT delays that were approved were also looked at since a higher proportion of approved TOT delays meant that an operator tended to request their TOT delays farther in advance. To investigate if the proportion of approved TOT requests varied across experimental conditions, a linear mixed model analysis was conducted, again with age as a covariant. Neither experimental factor was found to be significant (level of automation, $F(3, 14.8) = 0.60$, p = 0.623; level of re-planning, $F(1,14.7) = 0.204$, p = 0.658), but Figure 5-3 shows that there was a trend of increasing proportion of TOT delays approved as the automation level increased.

In summary, the main results from this section are the following:

- The number of TOT requests made were significantly higher for the high re-planning condition than for the low, and significantly higher for the active level than for all other automation levels.

- The proportion of TOT delay requests approved did not show significance but there is a non-significant trend of increasing proportion of TOT delays approved with automation level.
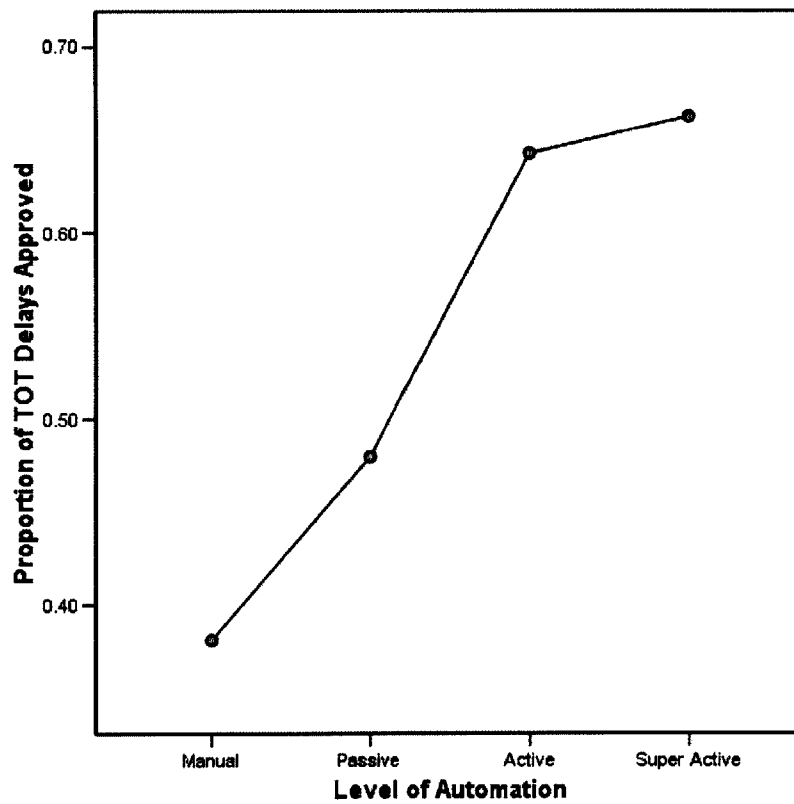

Figure 5-3: Proportion of Approved TOT Delay Requests Per Level of Automation

5.3 Wait Times

As discussed in Chapter 2, increases in system wait times are hypothesized to be a major source of performance decrements in time critical situations such as multiple

UAV supervisory control. System wait time occurred whenever any of the UAVs was in a degraded state (as defined in section 4.5.2) and needed attention from the operator. Overall system wait time (WTT) consists of interaction wait time (WTI), wait time in the queue (WTQ), and situation awareness wait time (WTSA). It should be noted that the sum of WTI, WTQ, and WTSA does not equal WTT if there is ever a period of time when two different types of wait times are occurring simultaneously (on a minimum of 2 vehicles), as the total system wait time metric does not double count these overlapping times. Wait times are a measure of system performance.

5.3.1   Total System Wait Time (WTT)

Total system wait time was computed as the aggregate time any part of the system was incurring wait time due to human-vehicle interaction. With this measure, it did not matter if a single UAV was waiting or all four, or the type of accumulating wait time, just that some part of the system required attention from the operator to improve its performance to an acceptable level. Total system wait time was significant for level of re-planning ($F(1,14.9) = 27.63$, $p < 0.001$) and level of automation ($F(3,14.9) = 4.631$, $p = 0.018$). As the factor level variances were not homoscedastic and there were normality violations, a logarithmic transformation of the data was required. There were no significant interactions.

Post-hoc comparison between automation levels revealed that total system wait time was significantly less under the super active level of automation as compared to every other level (SA vs. M, $p = 0.022$; SA vs. A, $p = 0.053$; SA vs. P, $p = 0.078$). From Figure 5.4, it can be seen that total system wait time for the active level under high re-planning conditions had the highest overall wait times. This data point was found to be significantly higher than under super active ($p = 0.027$) but no different from the passive or manual levels.

94
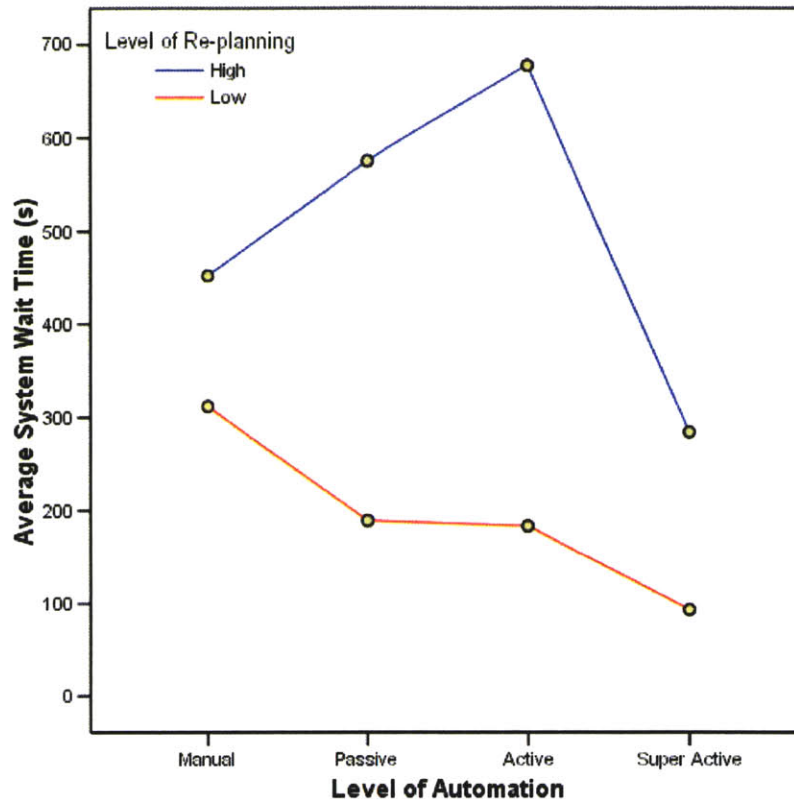
Figure 5-4: WTT Across Levels of Automation and Re-planning

5.3.2 Interaction Wait Time (WTI)

In the context of this experiment, interaction wait time was the subset of total human interaction time with the UAVs that took place while the vehicle under study was in a degraded state. WTI includes both the time operators spent actually entering commands to UAVs as well as the actual time spent planning the process. Examples of this are the time an operator spent planning a UAV's path out of a threat area, or the time a UAV spent loitering at a target, ready to fire before the operator communicated a firing command to it.

Figure 5-5 shows the marginal means for each of the experimental conditions. Interaction wait time was significant for level of automation ($F(3,13) = 8.08$, p = 0.003) and not significant for level of re-planning ($F(1,12.1) = 2.13$, p = 0.170). A logarithmic transformation of the data was required to meet equal variance and

95

normality assumptions, and there were no significant interactions between the factor levels. Post-hoc comparisons between automation levels showed that interaction wait time was significantly less for the super active and active levels than the manual and passive levels (SA vs. M, p = 0.001; SA vs. P, p = 0.004; A vs. M, p = 0.002; A vs. P, p = 0.006). From these results, two homogeneous subsets of automation levels were found: super active/active and passive/manual, meaning that super active and active LOAs produced statistically the same wait times, as did passive and manual.
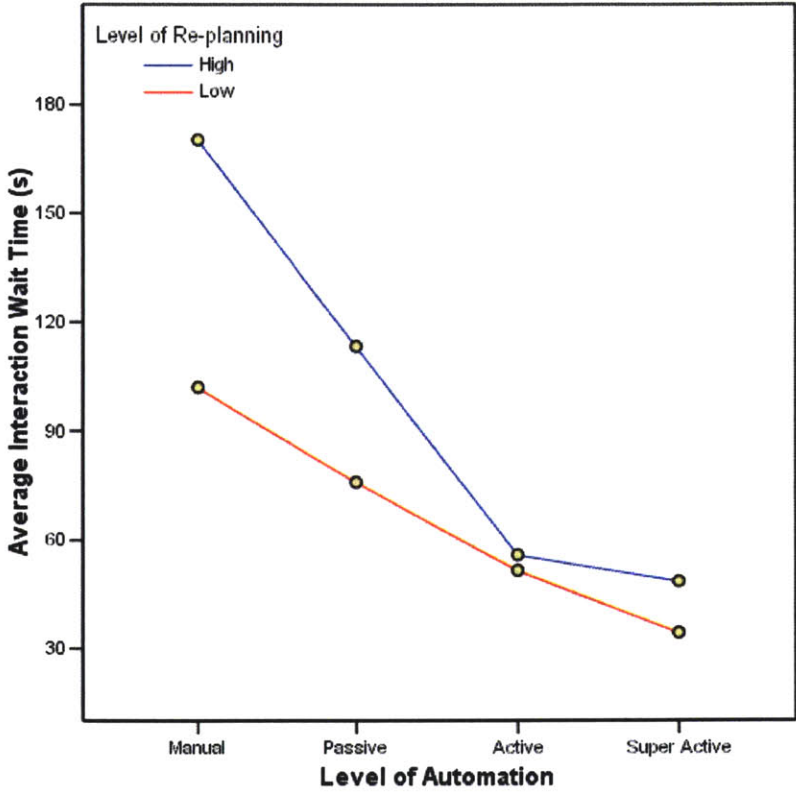

Figure 5-5: WTI Across Levels of Automation and Re-planning

5.3.3 Wait Time in the Queue (WTQ)

Wait time spent in the queue occurred in this experiment whenever two or more vehicles required operator attention simultaneously, and the operator moved immediately to the vehicle(s) waiting in the queue after finishing a task with another vehicle.

96

The amount of wait time spent in the queue was significant for level of re-planning $(F(1,12.3) = 18.08, p = 0.001)$ and marginally significant for level of automation $(F(3,13.2) = 3.10, p = 0.063)$ at the $\alpha = 0.05$ level. A square root transformation of the data with a 0.5 correction due to zero values was required to meet homogeneity and normality assumptions, and there were no significant interactions. The unusual spike in WTQ for the active level of automation under high re-planning in Figure 5-6 warranted further investigation. Post-hoc analysis showed that the difference between super active and both the active and manual automation levels' aggregate WTQ was significant (SA vs. M, p = 0.041; SA vs. A, p = 0.034). Further investigation showed that for the high re-planning condition, the active level had significantly more WTQ than both the super active and passive levels (A vs. SA, p = 0.009; A vs. P, p = 0.074), but the same as the manual level.
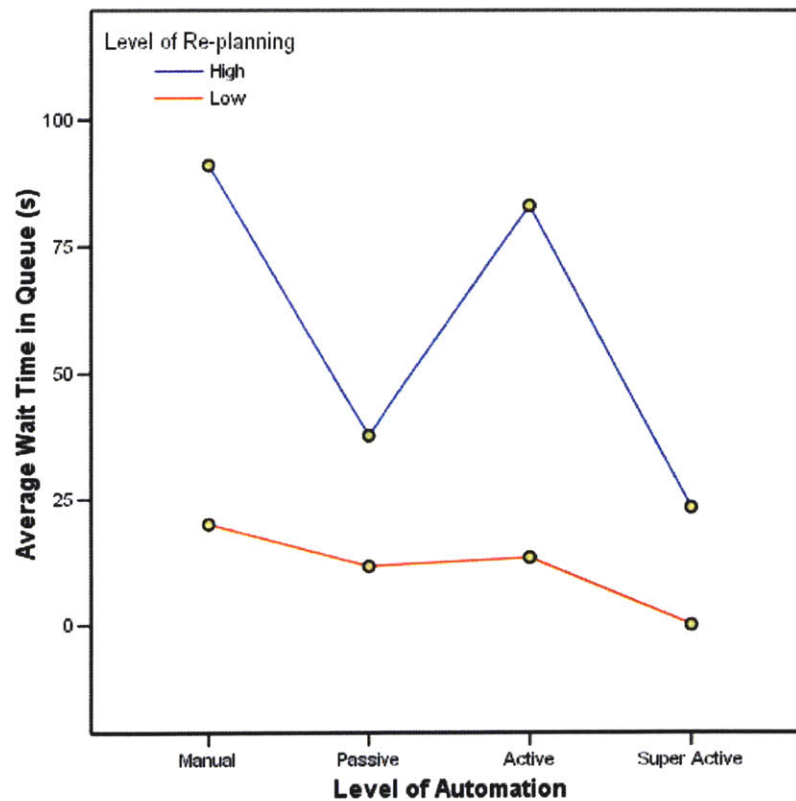

Figure 5-6: WTQ Across Levels of Automation and Re-planning

97

5.3.4 Situation Awareness Wait Time (WTSA)

Situation awareness wait time in the experiment was accumulated when at least one vehicle required attention but the operator didn't realize it. Common situations where WTSA was incurred included when subjects forgot to arm and fire on a target and left a UAV loitering unnecessarily at a target, or when a subject flew a UAV into a threat area. Situation awareness wait time was significant for level of re-planning $(F(1,12.3) = 18.70, p = 0.001)$ but not for level of automation $(F(3,13.2) = 2.14, p = 0.144)$. A square root transformation of the data was required to meet equal variance and normality assumptions, and there were no significant interactions. Despite the non-significant result for level of automation, exploration of individual cell means was performed to investigate the high WTSA levels of the passive and active levels under high re-planning, as can be seen in Figure 5-7. Post-hoc analysis showed that the difference in WTSA between the super active and both the active and manual automation levels was significant (SA vs. M, p = 0.055; SA vs. A, p = 0.063). For only the high re-planning condition, the active level was found to have significantly higher WTSA than the super active level (p = 0.046), and marginal significance was achieved between the active and manual levels (p = 0.091). This means that subjects in the active, high re-planning condition had significantly higher WTSA than the manual and super active levels for that particular test session. Interestingly, under manual automation there was no significant difference in WTSA between the high and low re-planning conditions.

5.3.5 Wait Time Proportions

As can be seen from Figure 5-8, total system wait time was dominated by WTSA regardless of scenario difficulty. The proportion of WTSA was not significantly different between re-planning conditions $(F(1,14.5) = 2.79, p = 0.116)$, but it made up 63% of all wait time in the low re-planning condition and 72% in the high re-planning condition, showing a trend of increasing proportion of WTSA with difficulty. The percentage of WTQ more than doubled from 5 to 11% from the low to high re-planning
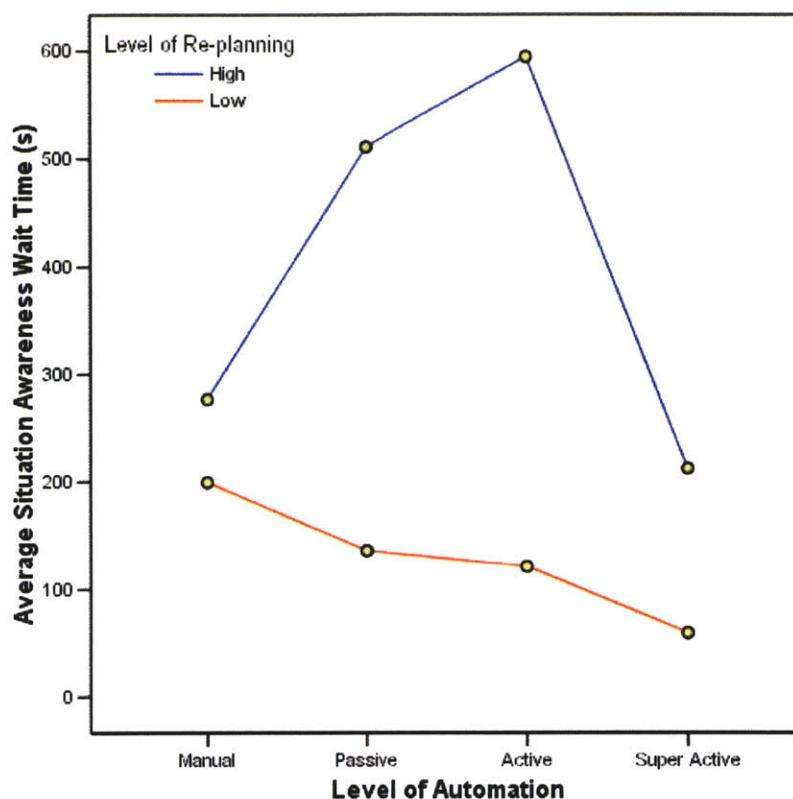
Figure 5-7: WTSA Across Levels of Automation and Re-planning

conditions, which was found to be a statistically significant difference ($F(1,12.6) =$ 4.85, p = 0.047). This was expected, as the high re-planning scenario contained a greater number of events, including instances of simultaneous re-planning events. The proportion of WTI followed the opposite trend, as it was significantly lower in the high re-planning condition ($F(1,10.2) = 9.3$, p = 0.012). The proportion of WTI was significantly lower in the high re-planning condition because it was not significantly different across re-planning conditions (Section 5.3.2), while total WTQ and WTSA (time, not proportion) were significantly larger in the high re-planning scenario (Sections 5.3.3, 5.3.4). Since WTI stayed approximately the same and the other types of wait times increased in the high re-planning scenario, the proportion of WTI decreased. There was no significant difference in the relative proportions of WTI, WTQ, WTSA or WTT across the different levels of automation

99

(a) Low   (b) High
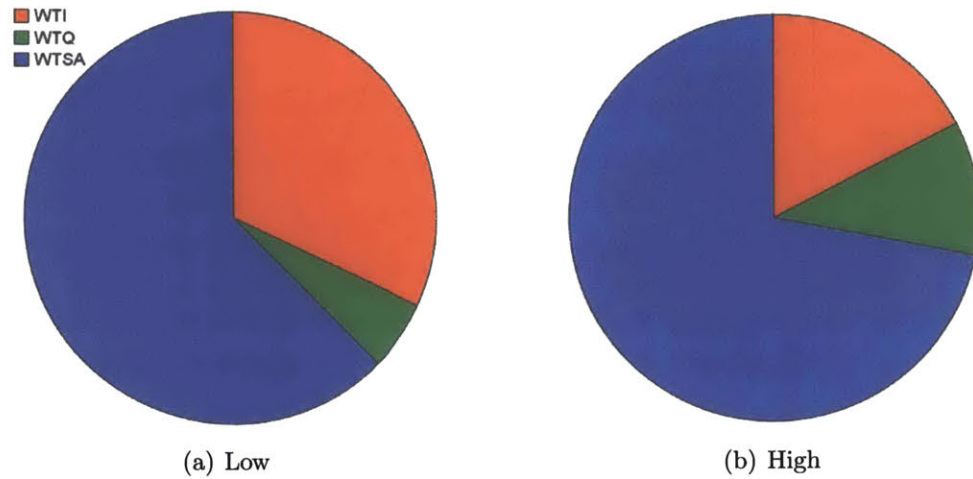
Figure 5-8: Proportions of WTI, WTQ, and WTSA Per Level of Re-planning

5.3.6 Correlation of Wait Times to Performance

Table 5.1 summarizes the correlation coefficients of the various wait times with the performance score results from Section 5.2.1. As can be seen, situation awareness wait time was highly negatively correlated with performance, while wait time in the queue had a moderately high negative correlation. While this does not necessarily prove high wait times caused lower performance, there is definitely a strong relationship between the two measures.

Wait Time	Correlation Coefficent, r	p-value
WTI	-0.255	0.230
WTQ	-0.593	0.002
WTSA	-0.894	<0.001

Table 5.1: Correlation Coefficients between Wait Times and Performance Score

5.4 Situation Awareness

In addition to human and system performance, a third major consideration in the evaluation of all HSC systems is situation awareness, which may not always correlate with other performance metrics. The objective SA data taken from online questions asked during test sessions for this experiment were not found to be consistent with

the subjective SA scale outlined in section 4.5.2. Therefore, as the objective SA measure's validity in the context of this study was questionable and since subjective SA scales based upon expert observer ratings have been found to be a reliable and valid measure way to measure SA [2], the objective SA scale was dropped in favor of the subjective one. Using this scale, differences in overall situation awareness were found to be significant for the two levels of re-planning ($F(1,11.7) = 15.75$, $p = 0.002$) and not significant for level of automation ($F(1,11.7) = 2.44$, $p = 0.112$). There was no significant interaction between factors.

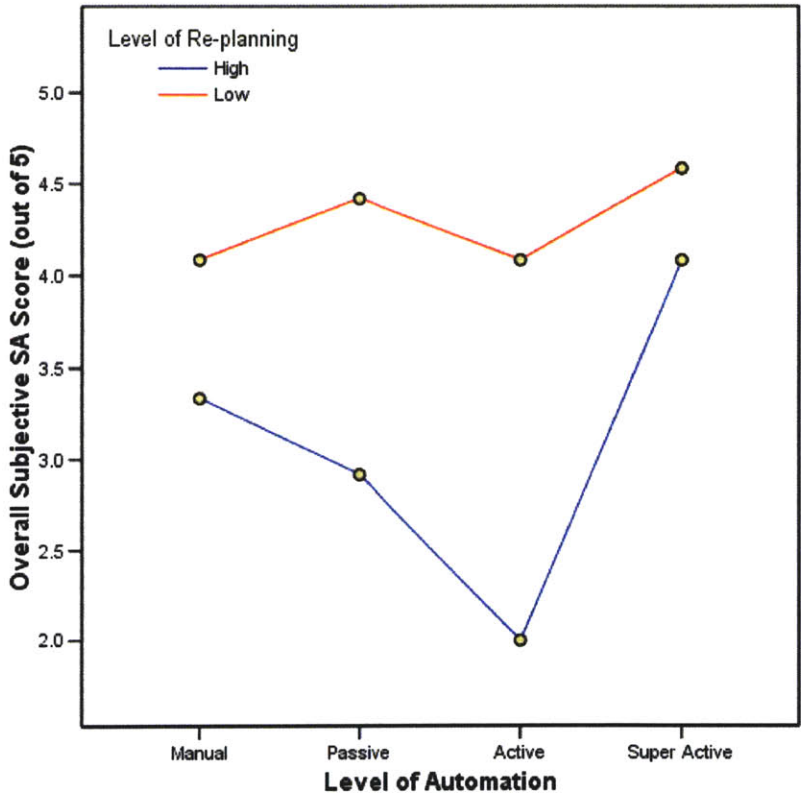Figure 5-9: Overall SA Scores Across Levels of Automation and Re-planning

From the marginal means plot (Figure 5-9), it can be seen that subjects using active automation had an unusually low subjective SA score in the high re-planning condition, so further investigation of this level of automation was conducted. The difference in SA between the active and super active automation levels under high

re-planning was significant (p = 0.034), but not significant between the passive and super active levels.

The overall subjective SA scores were then separated into level 2 SA and level 3 SA ratings and a non-parametric analysis was conducted on both datasets. Level 2 SA measures how well subjects comprehended the current situation. A Wilcoxon Signed Rank Test on the level 2 SA scores showed there to be a marginally significant difference between levels of re-planning (p = 0.065) and a Kruskal-Wallis test showed there to be a marginally significant difference between automation levels (p = 0.084). Figure 5-10 shows the means across each of the automation levels, which suggests that operators using the super active level had higher level 2 SA. Wilcoxon Rank Sum tests showed these differences to be significant (SA vs. M, p = 0.036; SA vs. P, p = 0.074; SA vs. A, p = 0.020).

Level 3 SA measures subjects' ability to project the current situation into the future. A Wilcoxon Signed Rank test on the level 3 SA scores showed that subjects' level 3 SA for the high re-planning scenario was significantly lower (p = 0.004) than for the low workload condition, but there was no effect of level of automation. This implies that the differences in overall SA were primarily driven by the significantly higher level 2 SA scores for the super active level.

5.5 Critical Events

As outlined in section 4.5.2, a critical event in the context of this experiment occurred whenever an operator committed an error of commission by erroneously firing upon a target that had already been removed from the ATO. The operator was notified of the ATO change through a high priority (red) message from intelligence in the chat box, which was additionally accompanied by an audio alert. There were nominally 3 opportunities for errors of commission in the high re-planning scenario and 2 in the low re-planning scenario. However, the number of potential opportunities for error varied by participant, as it was contingent on each operator's performance.

A non-parametric analysis was conducted to examine the effects of re-planning and
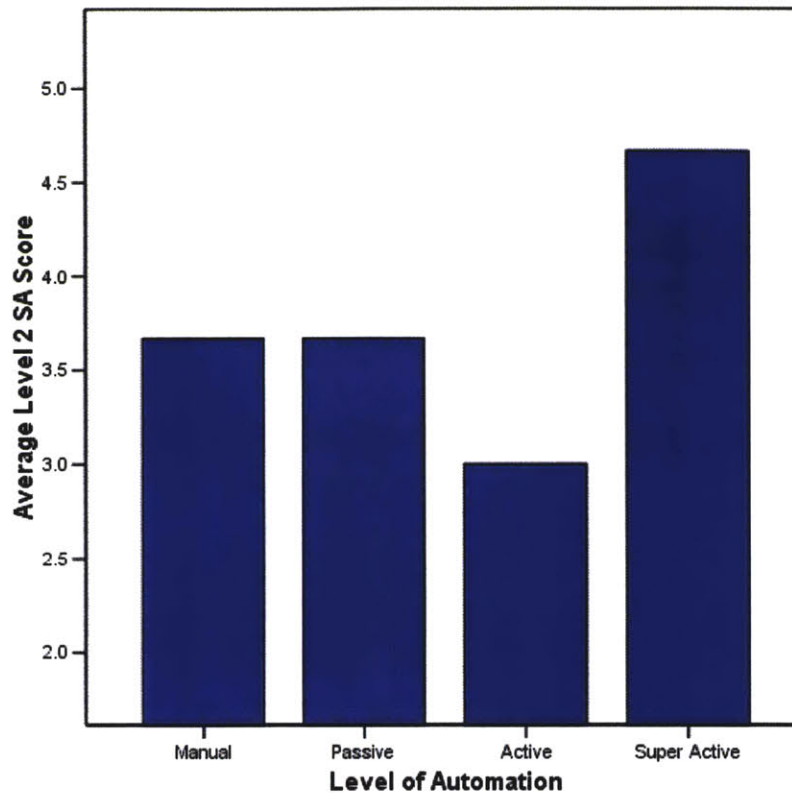
Figure 5-10: Average Level 2 SA Scores Per Level of Automation

level of automation on the number of critical events that occurred in the test sessions.
A Kruskal-Wallis test showed that there was no significant difference in errors of
commission between levels of automation, while a Wilcoxon Signed Rank test showed
there to be a significantly higher number of targets erroneously destroyed under the
high re-planning condition (p = 0.034). In fact, there was only one error of commission
under the low re-planning condition and 7 in the high re-planning scenario. Further of
investigation of this revealed that under high re-planning, subjects in the manual and
super active levels of automation committed more errors of commission than those
in the passive and active levels (Figure 5-11). Due to the low numbers of counts, a
Kruskal-Wallis statistical test comparing these experimental conditions to one another
did not prove significant, but the trend in Figure 5-11 is suggestive that operators
made more critical events errors in manual and super active levels of control.
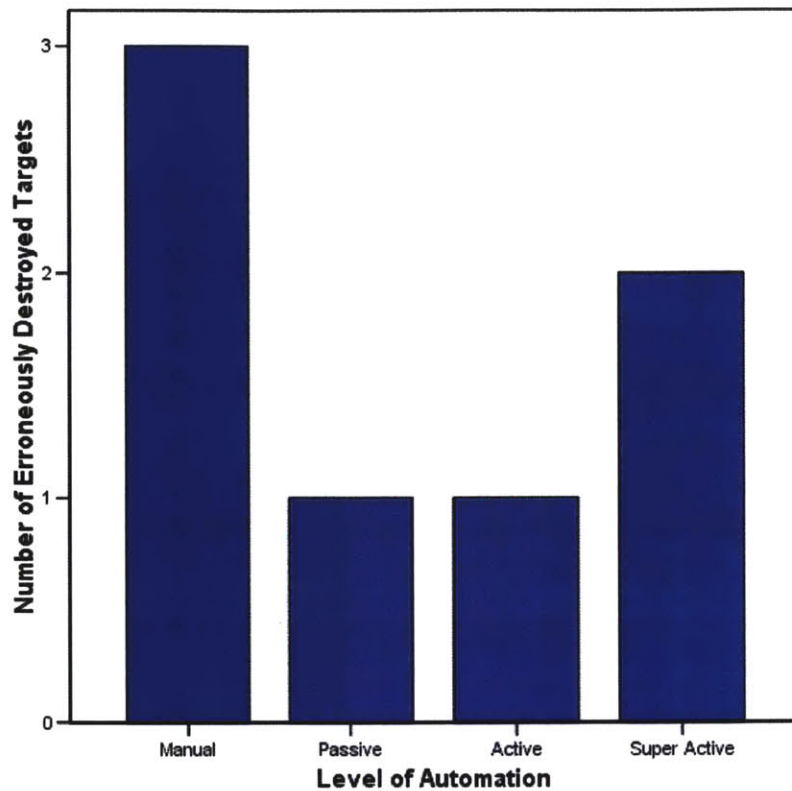
Figure 5-11: Erroneously Destroyed Targets Per LOA, High Re-planning Session

5.6 Workload Measures

Subjective workload measurements are thought to be useful measures of human performance, but they are prone to problems such as individual interpretations of rating scales. Tsang and Wilson [40] claim that subjective measures are reliable when used in conjunction with other objective measures. Thus, in this experiment workload was measured both objectively through response times to online questions, and subjectively through a modified NASA TLX survey administered after each of the test sessions.

5.6.1 Subjective Workload

Subjective workload was significantly different across re-planning conditions (F(1, 14.9) = 24.16, p < 0.001) but not across automation levels. From Figure 5-12, the unusually high subjective workload score of manual under low re-planning and active under high re-planning were identified as points requiring further investigation. Comparisons within automation levels across re-planning conditions found that subjective workload was significantly greater in the high re-planning scenario for every level of automation except for manual, where no significant difference was found. Comparison of the manual level to all other LOAs in the low re-planning scenario was not significant (p = 0.093) and the active level to all other LOAs in the high re-planning scenario was not significant either (p = 0.126), though the trend from Figure 5.12 is clear and significance was nearly achieved.
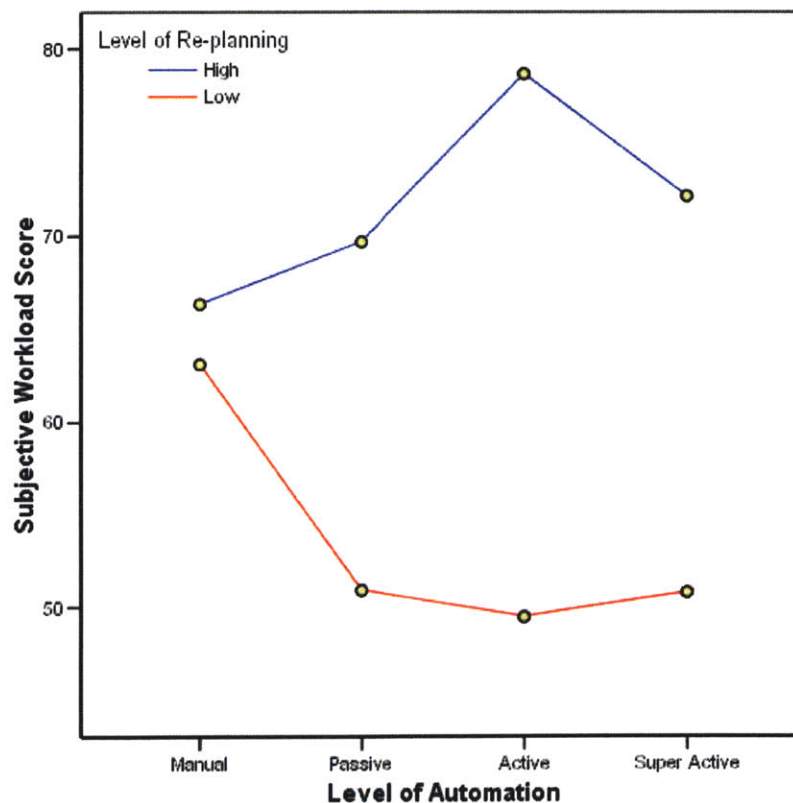
Figure 5-12: Subjective Workload Across Levels of Automation and Re-planning

5.6.2 Secondary Workload

As can be seen from Figure 5-13, secondary workload was significantly higher (and thus spare mental capacity significantly less) for the high level of re-planning ($F(1,11.5)$ = 14.69, p = 0.003), and significant differences across the levels of automation were found as well ($F(3,12.6)$ = 9.22, p = 0.002). There were no significant interactions. Cell comparisons showed that the manual level had significantly less spare capacity than the passive or active levels (M vs. P, p = 0.001; M vs. A, p = 0.017), but there was no difference between the manual and super active levels. However, under the low re-planning condition, the manual level had significantly higher secondary work-load than all other LOAs including super active (p = 0.061). Comparisons across re-planning conditions within the passive and manual levels also showed no signifi-cant difference for both levels (p = 0.299, passive; p = 0.089, manual), meaning the spare capacity of the manual and passive levels did not change with re-planning level.
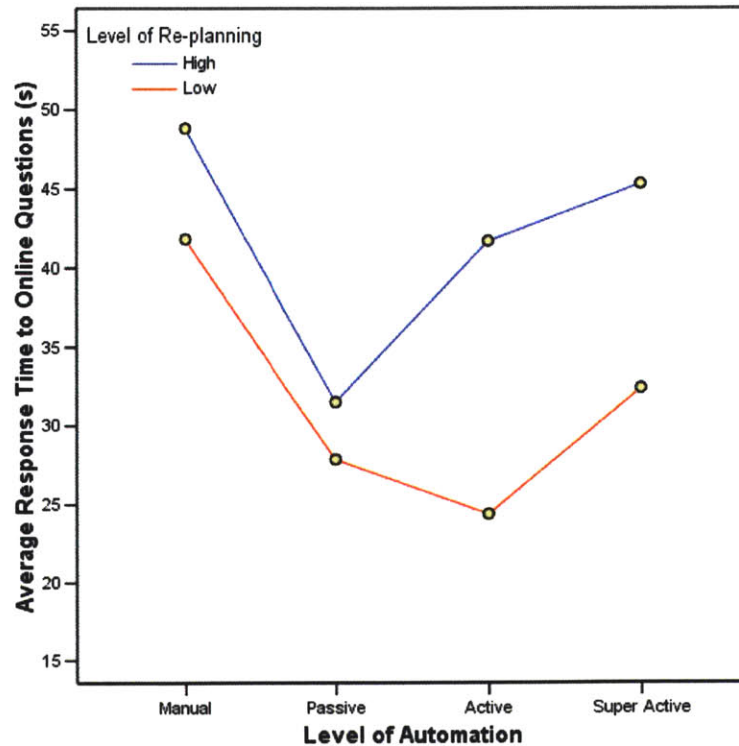
Figure 5-13: Average Response Times Across Levels of Automation and Re-planning

5.7 Summary

Table 5.2 (next page) summarizes the main effects results for all dependent variables presented in this chapter. In general, the level of re-planning was significant across all dependent variables (WTI excluded), while the p-values of many of the tests involving the level of automation factor were only marginally significant. However, individual cell-to-cell contrasts in such cases often revealed interesting relationships between the different levels of automation. Table 5.3 (next page) summarizes these contrasts in terms of which levels of automation were the best and worst performers for each dependent measure. In cases where levels of automation could not be differentiated from another, both levels are listed as having equally poor or good performance. For a summary of the descriptive statistics for all dependent measures, see Appendix C.

Dependent Variable	Level of Automation	Level of Re-planning
Performance Score	0.076	0.001
Number TOT Delay Requests	0.096	0.059
Proportion of TOT Requests Approved	0.623	0.658
Interaction Wait Time (WTI)	0.003	0.170
Wait Time in the Queue (WTQ)	0.063	0.001
Situation Awareness Wait Time (WTSA)	0.144	0.001
Situation Awareness	0.112	0.002
Erroneous Target Critical Events	0.878	0.034
Subjective Workload	0.779	<0.001
Secondary Workload	0.002	0.003

Table 5.2: Summary of Main Effects (p-values)

Dependent Variable	Worst Performer	Best Performer
Performance Score	A	SA
Number TOT Delay Requests	A	M
Proportion of TOT Requests Approved	M	SA/A
Interaction Wait Time (WTI)	M/P	SA/A
Wait Time in the Queue (WTQ)	A/M	SA/P
Situation Awareness Wait Time (WTSA)	A	SA
Total System Wait Time (WTT)	A	SA
Situation Awareness	A	SA
Erroneous Target Critical Events	SA/M	A/P
Subjective Workload	A/M	SA/P
Secondary Workload	M	P

Table 5.3: Best and Worst Levels of Automation for Dependent Measures

Chapter 6

Discussion

This chapter discusses the results presented in Chapter 5, how they relate to previous work, and comparisons are made to the hypotheses proposed in Chapter 4. Four main areas are addressed: workload, human and system performance, situation awareness, and cognitive strategies.

6.1 Workload

In general, all dependent variables of this experiment changed in predictable ways across the two levels of re-planning, which represent high and low levels of workload. As workload increased, performance and situation awareness decreased, and all types of wait times, occurrences of critical events and subjective measures of workload increased. As expected, spare mental capacity measured through secondary tasking decreased with increasing workload. These results demonstrate that task performance can vary dramatically under different workload conditions, and one design condition that may effectively support a task under a low workload situation may be counterproductive in a high workload setting.

Two specific trends in the workload score were noted that deserve further discussion: 1) the manual level showed no increase in perceived workload or decrease in spare capacity across re-planning conditions, and 2) the super active had the same spare capacity as the manual level, which was higher than the active or passive lev-

els. The first result implies that operators under the manual level perceived they were under high workload irregardless of session difficulty, and that this was in fact true because there was no change in spare capacity either. Without access to a visual representation of the future mission plan (the timeline), operators in the manual level were not able to easily perceive changes to the mission plan, so even under low re-planning they were constantly scanning the ATO table for changes.

The response times to online questions, which were a measure of secondary workload, showed that the manual and super active levels of automation had the same average response times, which were greater than the passive and active levels. Therefore, subjects in the manual and super active levels had less spare mental capacity. This is an interesting result for the super active level of automation, because this level was supposed to have the lowest level of workload, which is related to spare capacity. However, it is likely that that even though super active operators had good SA, they were not directly involved with mission execution actions so they had to work mentally harder to stay "in the loop" and maintain those high levels of SA.

6.2 Performance

6.2.1 Human Performance

The significant difference in performance scores across re-planning levels in this experiment indicates that the number of re-planning events is an important influence on command and control mission difficulty, as expected. Under low levels of re-planning, subjects using all levels of automation performed equally well, but under high re-planning subjects using active automation had lower performance than all other automation levels. The poor performance of the active level is surprising, as Ruff et al. [35] found that a medium level of automation called management-by-consent corresponding approximately to the active level in this study had performance advantages over manual and management-by-exception (super active) conditions for multiple UAV supervision. Many previous studies from other domains have also found

110

collaborative types of automation to have superior performance to highly automated or manual systems [18, 21, 42]. However, a primary limitation in such systems as demonstrated in this study is that users can often be overwhelmed by the large array of possible actions they can take. This can be particularly problematic for command and control systems under temporal constraints and with significant uncertainty, as was the case in this study.

The performance decrement under the active level of automation can be attributed to subjects' inappropriate use of the "Request TOT" function in the MAUVE simulation. Subjects often did not generate appropriate stopping rules when trying to achieve a particular schedule move. At the detriment of other tasks and vehicles requiring their attention, subjects often focused on obtaining a particular delay they wanted until they obtained it. Operators' behavior also suggested they believed they could always make up for lost time through this function, when in reality the probability of approval became unreasonably low unless subjects were able to plan at least several minutes in advance. Rather than cutting their losses and choosing to give up on a target to improve a UAV's arrival time at subsequent targets, subjects often tried until the very last possible instant to obtain the TOT delay they needed. This was likely due to misjudgment of the probability of obtaining a last minute approval, as humans are not good estimators of chance and typically overestimate very small probabilities [41].

Subjects' second misuse of the TOT delay function was failing to consider the impact of using it on the arrival times to subsequent targets in the queue, as movement in one part of the schedule almost always affected ETAs to other targets. The ability to move a target's TOT was an intervention meant to be used sparingly, and the consequences of doing so were rarely considered by operators. TOT requests were primarily driven by immediate, short-term needs and they tended to cause problems for subjects later in the schedule, especially if they were not noticed until much later. Given the time-critical, demanding task, it was too difficult for operators to execute and re-plan according to the ATO as well as consider schedule optimizations far into the future.

The reverse shading technique used to highlight areas of potential high workload for the active automation was likely was the cause of the overuse of the TOT delay request, as it drew operator attention to specific areas of the timeline farther into the future. Giving operators information about potential problem areas far in advance seemed to change their strategy from locally optimizing to that of globally optimizing. Given that at any time there could be many different areas of high workload on the timeline at once, and the tight constraints on the deadlines operators were under, this was a very difficult task to achieve, often overwhelming operators and resulting in them creating a schedule worse than before. Due to its complex nature, this process also took a long time to perform and often caused other performance problems by taking the operator's focus off of the immediate, near term needs of the UAVs (local optimization). However, significant performance and situation awareness decreases due to these attempts at global optimization really only showed up in the high re-planning scenario when there was much less room for operator error and inappropriate attention allocation.

In contrast to the active level, it was observed that users in the manual level never used the TOT delay button to relieve high workload; they only used it for short-term movement of deadlines they were going to miss. This was because identifying areas of high workload involved comparisons between multiple vehicles, which were facilitated by the vertically stacked timelines, while simple schedule changes for a single UAV did not require as much data fusion. Users of the manual level of automation were pure local optimizers, and it served them well as they had an average level of human performance as compared to the other automation levels, which was higher than expected.

In general, the visual timeline format seemed to enable users to more successfully identify problematic, high workload sections of the schedule more easily and farther in advance, as there was an increasing trend (non-significant) of proportion of approved TOT delays with automation level. While highlighting these potential problem areas turned out to be detrimental to performance for the active level, this knowledge was beneficial to schedule planning so long as operators did not feel compelled to globally

optimize like they did in the active level. The lower number of TOT delays requested for the passive and super active levels as well as observed behavior in the test sessions confirms that this was the case.

6.2.2 System Performance

Interaction Wait Time (WTI)

Figure 6-1 illustrates how the predictions for the trends of WTI with levels of automation outlined in Section 4.1 compared to the results obtained in the experiment. In general, a decreasing trend of WTI with increasing levels of automation was found, with the super active and active levels of automation having significantly less WTI than the passive and manual levels, which was generally consistent with expectations.
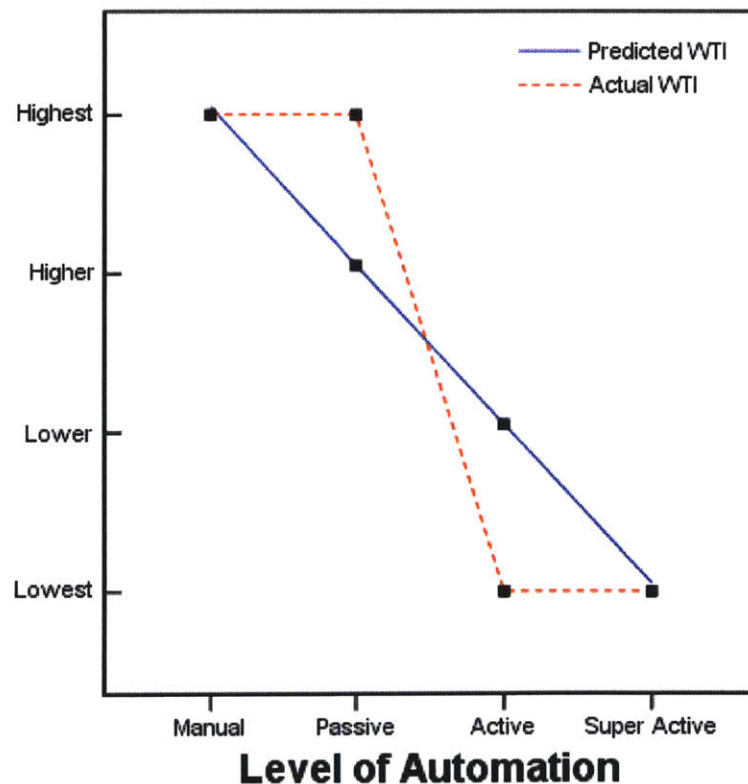

Figure 6-1: Comparison of Predicted and Actual WTI Trends

Wait Time in the Queue (WTQ)

The manual level of automation had statistically the same level of WTQ as the active level, while the super active had significantly lower WTQ than both of these levels. The passive level fell somewhere in between these sets of automation levels, not significantly different from either one. These results approximately followed the predicted trend of decreasing WTQ with automation level made in Chapter 4, except for the relatively high average WTQ for active automation (Figure 6-2). Quantitative analysis of WTQ accumulation in the active, high re-planning test sessions showed that the majority was accumulated in several large queues that formed late in the scenarios when multiple, difficult re-plans were required of the operator. While the accumulation of WTQ in this way was not unique among automation levels, the queues did build higher in the active level, which is why WTQ is significantly larger in this condition. This is not surprising given the extra time operators were spending attempting to adjust TOTs.

Situation Awareness Wait Time (WTSA)

Average system WTSA did not closely follow the predicted trend across the levels of automation, as can be seen from Figure 6-3. Users of the super active level of automation had the best WTSA times, while active had the worst and manual/passive were in between. As WTSA is dependent on situation awareness levels and followed the same trend, detailed discussion on why these trends were observed is left until section 6.3.

WTSA increased with along with level of re-planning, as was expected. Higher WTSA was expected with greater workload because subjects were already under high workload even in the low re-planning condition, so adding more workload on top of that was expected to just decrease their performance and lower their SA even more. As already discussed, lower SA means, in general, higher WTSA.
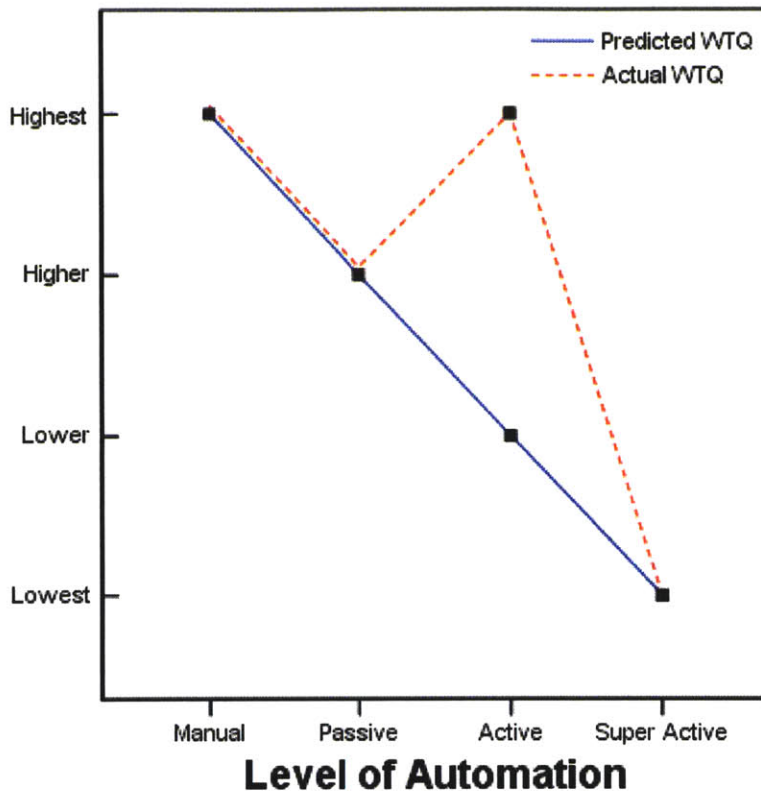
Figure 6-2: Comparison of Predicted and Actual WTQ Trends

6.2.3 Performance Metrics Conclusion

One of the most important findings from this study was that total system wait time was dominated by wait times caused by loss of situation awareness, rather than interaction or queuing times. It is likely that several factors including additional training, a more refined interface, and additional subject motivation would decrease the proportion of total wait time from WTSA, but even a significant decrease would still leave WTSA as the largest contributor to system wait times. Despite the fact that WTQ and WTI occurred more often, WTSA was the largest contributor to WTT because the average delay due to WTSA was many times that of WTI, and to a lesser extent, WTQ.

Even though subjects performed a relatively complex task in the MAUVE simulation, compared to an actual command and control situation, the task was simplified.
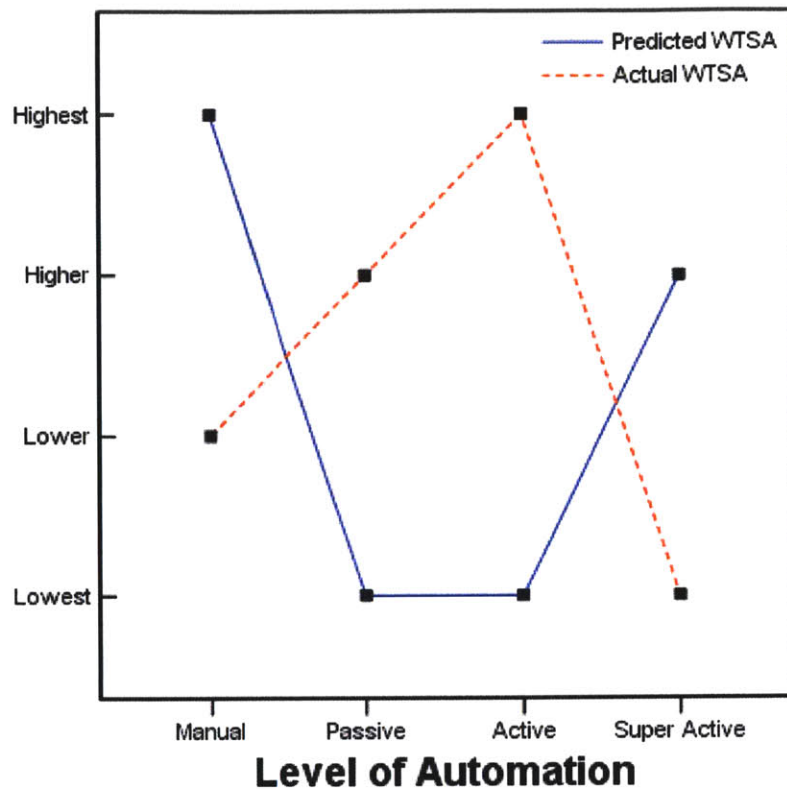
Figure 6-3: Comparison of Predicted and Actual WTSA Trends

However, it is likely all wait times will increasingly scale proportionally. The effect of re-planning on wait time proportions in this experiment shows that at the very least, the proportion of WTSA will not decrease as task complexity and difficulty increase. As has been shown in this experiment, increasing vehicle autonomy generally decreases WTI and WTQ, but not necessarily WTSA. This is because WTSA is never certain to occur and can only be probabilistically modeled based upon a subject's level of SA, number of vehicles to be controlled, etc. As system and situation awareness wait times were shown to be highly negatively correlated with performance in Section 5.3.6, this implies that there is an inherent limit to the number of vehicles any human can effectively supervise, no matter the degree of autonomy provided to them.

6.3 Situation Awareness

The subjective SA ratings and WTSA measure were consistent with each other in that under high workload, operators had better SA in the super active automation level and a degraded level of SA in the active automation level. This was contrary to expectations, as the manual and super active levels of automation were expected to have lower SA levels than medium levels of automation. This was predicted because an operator's level of SA can decrease under high workload due to competition for attentional resources, but also decrease under low workload due to boredom and complacency.

The active level of automation had the worst level of SA because of operators' need to globally optimize their schedules as detailed in Section 6.1. This decreased their level of SA for several reasons: 1) In seeking to globally optimize, subjects were actually narrowly focused on the timeline display and not the map display, and 2) the optimization itself was a very high workload process, causing operators to become overloaded.

The compelling nature of the timeline decision support likely induced lower SA for the active level, and possibly for the passive level as well. It was observed in test sessions that many subjects in the active and passive levels exclusively looked at the timeline decision support to schedule their time, and rarely looked at the map display. Across all levels of automation, subjects' general strategy was to decompose the complex time management task into more manageable chunks. Under the active and passive levels, a visual cue subjects used extensively to allocate their attention was to look for yellow or red areas on the timeline (the arming and firing windows) near the thick black line on the left side of the decision support, which told subjects an arming and/or firing window would occur in the near future. It was observed that subjects' performance in these levels was comparable to the manual or super active levels in manipulating the schedule such that the UAVs arrived on time to each of their targets. However, operators with passive and active tended to ignore the threats on the map display and in many cases ran their UAVs straight through them, thus exhibiting

attentional blindness and cognitive tunneling. Another common instantiation of poor relational SA under the passive and active levels occurred when subjects tried to fire on an expired target because the timeline told them a target TOT was occurring, only to find the relevant UAV loitering at the entirely wrong target. This generally happened because subjects hadn't bothered to verify the UAV's location on the map display in quite some time because they were overly focused on the timeline.

In contrast, subjects under the manual level frequently switched their attention between the two displays, as their version of the decision support showed absolute time, requiring them to constantly compare times on the decision support to the mission time clock on the map display. It was noted during the test sessions that manual level subjects made frequent, visible head movements that indicated attention was being shared between the screens on a consistent basis.

Subjects in the super active level also generally did not have an attention allocation problem caused by the decision support. As the automation took care of arming and firing, subjects in this experimental condition did not rely on the timeline for scheduling these actions. Further evidence of subjects' lack of reliance on the decision support in the super active level was the low usage of the exception boxes that were used to cancel arming and firing when necessary. These were located to the right of the timeline decision support window (Figure 3-7). Of the 13 possible instances across all super active test sessions when arming or firing was correctly cancelled as part of a re-planning effort, the exception boxes were only used 5 times. This means that subjects were not just blindly clicking on the exception boxes whenever they saw message that a target had been removed from the ATO. Successful re-plans in these instances could be accomplished by either using the exception box or removing the target itself from a UAV's mission plan, which involved selecting that target and using the "Remove Target" function. Instead of relying on perception and overly trusting the automation, they commonly removed the relevant target from the mission plan, despite it taking more steps to achieve the same result. This indicates that subjects in the super active condition understood the problem at a deeper level, which required a high level of SA.

A break-down of the SA scores into level 2 and level 3 SA showed that there was no difference in level 3 SA across automation levels, but that super active automation had significantly higher level 2 SA scores. This is further evidence that super active operators were not overly focused on their decision support, as level 2 SA was highly dependent on how aware the operator was of relational information on the map display. This occurred because subjects in the super active level had fewer actions to perform, and as a consequence had more time to observe events on the screens. However, as a result being "out-of-the-loop", i.e. not always cognitively engaged with the vehicles, they gained no benefit for level 3 SA with this extra time.

Despite having the highest level of SA, driven by level 2 SA, subjects in the super active automation condition erroneously destroyed more targets than those under passive and active automation. This indicates that subjects in super active suffered from complacency, as they were content to let the computer take care of the arming and firing actions without closely monitoring them. This is interesting because it seems to indicate that users in super active trusted the computer for weapons release but not target removal. This may be a case of automation bias, as the subjects were reluctant to intervene in what they deemed was the automation's responsibility, arming and firing, while they had no problem re-planning routes because they thought that was the task that had been assigned specifically to them.

The manual level of automation had SA scores that were relatively higher than expected in this experiment, as they were predicted to be among the lowest of all automation levels. Considering both the subjective and secondary workload (spare capacity) scores, it can be concluded that as expected, the manual level of automation had the highest workload of all automation levels. However, SA was expected to be significantly lowered due to this high level of workload, which did not happen. Based on post-test feedback and observations in test sessions, it was observed that with experience, subjects established a coping strategy where they developed scan patterns much like that taught to pilots that enabled them to quickly detect when an off-nominal situation existed in the mission plan. Operators in higher levels of automation that had lower workload had less of a need to develop coping strategies,

119

so as a consequence, the operators in the manual condition likely were the most engaged in the scenario events. However, operators using the manual automation had a moderate level of SA, but actually had the worst critical event performance. Even though the scan pattern helped manual operators reduce their workload to manageable levels, the pattern itself was quite automatic and made operators more prone to making errors of commission as they focused on maintaining it. This result also could be a direct consequence of operators' high workload, as there was usually only a short time between the abort firing command being issued and the TOT for the affected target.

6.4 Cognitive Saturation and Coping Strategies

During testing, a measurable point of cognitive saturation was observed when subjects were overwhelmed, either from workload generated by frequent re-planning events, or a loss of SA. This saturation point was only seen in the high re-planning test sessions, as the low re-planning scenario had enough rest periods to allow operators to recover without serious loss of system performance. Once overloaded, operators' behavior followed two distinct patterns: 1) they either began to supervise fewer vehicles, ignoring one or more entirely, or 2) they attempted to supervise all of them, missing a majority of their deadlines while they tried to do so. Figures 6-4 and 6-5, which plot actual UAV paths against the optimal paths over the course of two high re-planning trials, illustrate both strategies.

In Figure 6-4, after target A emerged, the operator became confused as to the appropriate routing for UAV 2. The path of UAV 2 reflects the operator's loss of SA as he tried various different but incorrect routings over the next few minutes, before giving up on those vehicle's targets entirely. While doing this, the operator successfully followed the mission plan for UAVs 1 and 4, as can be seen by the close match between the optimal and actual paths for those vehicles. However, UAV 2 was left loitering at target B for over 7 minutes, causing it to miss all 3 remaining targets on its route. In this case, the operator employed a cognitive shedding strategy

whereby interaction effort was focused on a smaller subset of vehicles than originally tasked to the operator. This was able to reduce the operator's mental workload to manageable levels and so performance on the remaining vehicles was not significantly degraded.
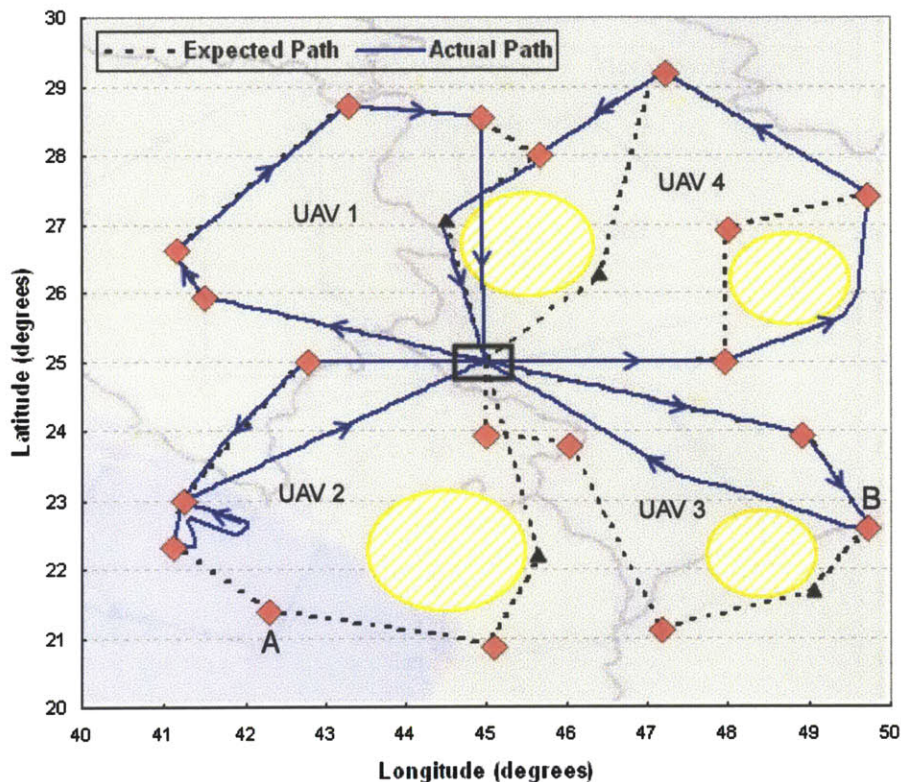

Figure 6-4: Coping Strategy 1 - Cognitive Shedding

In Figure 6-5, the operator's saturation point occurred when he left UAV 2 loitering at target A for nearly 5 minutes before realizing it had stopped executing the mission plan. The operator then tried to salvage some later targets on the route by heading to target B, but gave up half-way there before sending the UAV to base. In the intervening time, the operator also tried to control the other 3 UAVs, but left UAV 3 loitering at target C for several minutes, then routed it through the nearest threat area, missed target D with UAV 4, and also missed target E with UAV 1,

121

though this is not visible from Figure 6-5 because the UAV was guided to the target but did not fire on time. After giving up trying to hit any more targets with UAV 2, the subject's performance on the remaining 3 vehicles improved dramatically. In this case, the operator tried to cope with all 4 vehicles at once while in an overloaded state and ended up performing very poorly with all of them.
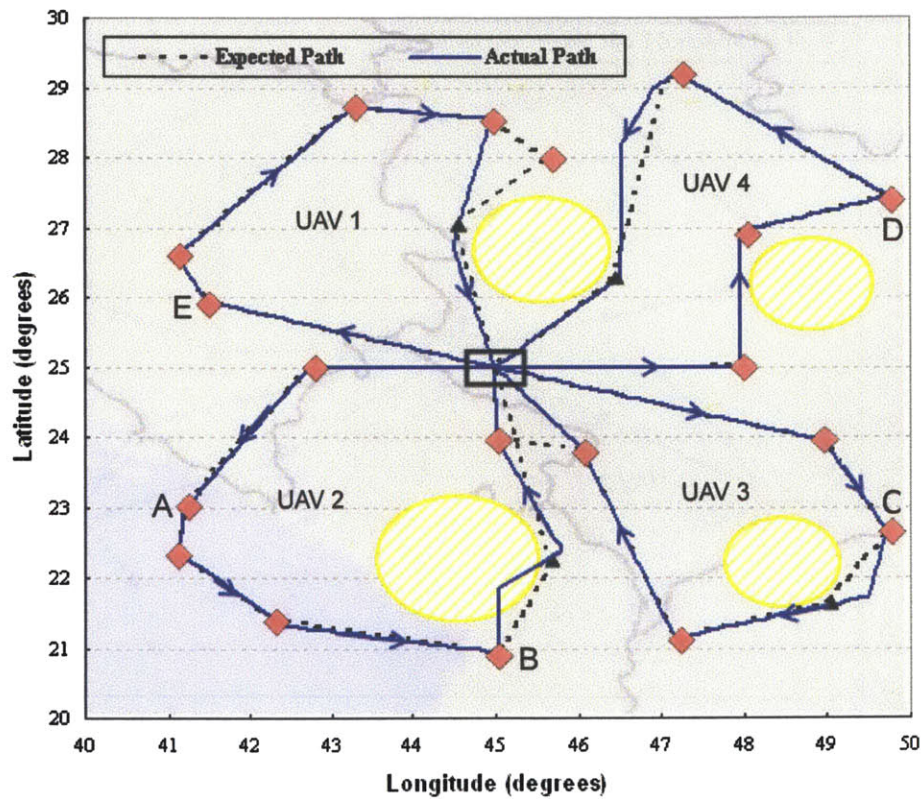

Figure 6-5: Coping Strategy 2 - Degraded Level of Management

Chapter 7

Conclusion

This chapter briefly summarizes the motivation for this research, presents the key findings of this study and makes recommendations for future work.

7.1 Overview of Study Motivation

There is increasing interest in the military to develop fleets or "swarms" of highly autonomous vehicles, but in order to make this possible, significant advances in vehicle control and other areas such as human-system interfaces will be required. As unmanned vehicles become more autonomous, the role of the human will shift from manual control to supervisory control. HSC tasks are primarily cognitive in nature and do not require constant attention, so this will allow operators to effectively supervise more than vehicle. In the future vision of an operator supervising multiple unmanned vehicles such as UAVs, it is desired to maximize the number that a single human can effectively supervise. The major limiting factor on this number is operator workload, which has clear upper limits. One way of mitigating operator workload and improving overall system performance is through automated decision support. However, it is not clear what type or level of decision support is appropriate for supervision of multiple vehicles, as very highly automated systems can induce low operator SA. This study sought to explore this problem further by investigating which level(s) of decision support could be appropriate for multiple vehicle supervision

in a time critical setting, both from a human performance and system performance standpoint.

7.2 Significant Findings

To this end, this study explored the human and system performance characteristics of four different types of decision support, termed manual (LOA 1), passive (LOA 2), active (LOA 4), and super active (LOA 6) for human supervision of 4 UAVs performing a suppression of enemy air defenses mission. The super active level was found to have the best performance characteristics in terms of human and system performance as defined by the performance score (Section 4.5.2) and wait times (Section 4.5.2). In addition, operators using the super active level exhibited high situation awareness for level 2 SA but not level 3 SA. This occurred because operators in this level of automation had more time to observe events on the map display, but they were not as cognitively engaged in the task as they had to perform less data processing. Thus, they were very effective at perceiving changes in the current state but no better at predicting future situations. However, operators under super active automation became complacent, allowing the automation to execute firing actions with too little supervision. As a result, operators failed to veto the arming/firing process several times and had a higher number of erroneously destroyed targets than the more moderate levels of automation, passive and active.

On the other end of the automation spectrum studies in this experiment, subjects using the manual level of automation performed better than expected. While subjects felt that they were always under high workloads despite the level of re-planning, they still managed to perform overall as well as subjects with the passive level of automation. However, the manual control strategy had several drawbacks, such as higher system wait times than all but the active level under high re-planning, and lower spare capacity so it is likely operators could not handle a much greater pace of operations or an increase in the number of vehicles supervised as easily as other automation levels. Moreover, most significantly the manual level had the worst critical

event performance, inadvertently firing on the highest number of unapproved targets.

The level of automation that produced the most surprising results was the active level of decision support, which involved a simple form of collaborative automation that provided predictions for possible periods of task overload as well as possible courses of action to relieve the high workload. It is important to note that subjects were aware that the predictions were not always accurate and that the further into the future, the less reliable the predictions. Collaborative automation is generally thought to be beneficial for performance as it allows humans both a better understanding of the system's operation as well as provides them with avenues to explore the possible solution space and generate satisficing and creative solutions [18, 21, 42]. However, this was not the case in this study, as users of the active level of automation had the worst performance of any automation level and the lowest situation awareness levels as well. This statistically significant performance degradation is attributable to an unintended consequence of the automation where the graphical visualization of the computer's predictions caused users to try to globally optimize the schedules for all UAVs instead of locally optimizing schedules in the immediate future. Attempting to globally optimize their schedules was a difficult task for operators because the time-critical nature of their mission meant that there was little room for error. Under the higher workload condition there were large decreases in performance and situation awareness as operators in the active condition quickly became overwhelmed and confused by their failed attempts at global optimization. This is an example of a hidden problem with collaborative human-computer decision making - the human can be overwhelmed by the choices they have, and therefore there are benefits to limiting the solution space.

This study found that the passive level likely has the greatest cost/performance benefits, as both the human and system performance was always at least at a moderate level, its users were under some of the lowest workloads of all the levels, and it tied for the lowest number of targets that were erroneously destroyed with the active level. Essentially, it was the only reliable performer as all other levels of automation had major drawbacks associated with them.

The wait time model proposed in this study to predict the relative sizes of wait time components across levels of automation was a good predictor of WTI and WTQ, but not WTSA. The reason for this is because when the model was formulated, it was assumed that the passive and active levels would have moderate levels of workload while the super active would have very low and manual very high workload. As an operator's level of SA can decrease under high workload due to competition for attentional resources [43], but also decrease under low workload due to boredom and complacency [31], moderate levels of workload result in optimum performance. In reality, all operators were under significant workload and any amount of workload relief provided performance benefits as well as improved SA. As a consequence, the super active level had the lowest total of WTSA in this study. Active level's very high WTSA times also were not predicted by the proposed model, and deserves more investigation.

A significant finding from this study was that total system wait time was dominated by wait time caused by a lack of situation awareness. This was consistent across all levels of automation as well as levels of re-planning, so it is likely to be generalizable to a larger set of unmanned vehicle missions and applications. This has implications for multiple vehicle control, as wait time in the queue and interaction wait time can be decreased with greater levels of autonomy, but wait time due to lack of situation awareness cannot be entirely eliminated. This shows that no matter how much autonomy is built into unmanned vehicles, there will be a clear upper limit on the number of vehicles that any one person can supervise because of the need to stay cognitively aware of unfolding events.

7.3   Recommendations and Future Work

The following are recommendations for future follow-on work based on this thesis:

- This work did not attempt to quantify the cost of wait times, as described by Freed et al. [15]. However, this is an important area of research to pursue because wait times in themselves do not cause performance problems; it is the

126

context in which they occur. For example, it is often the case that there is no measurable cost to wait time until a discrete failure occurs, at which time a large, fixed cost is incurred such as in the case of dropping a bomb on a wrong target. This event likely has a much higher cost than missing targets or battle damage assessment assignments.

- The results from this experiment could be extended to fan out and predictions made as to the theoretical number of "maximum" vehicles an operator could handle in the MAUVE simulation with the different levels of automation. It would be of particular interest to compare the results using the original equation for fan out (Equation 2.1) proposed by Goodrich et al. [17] and the fan out equation presented in this thesis incorporating wait times (Equation 2.3).

- A focus on the active, or collaborative level of automation in this study should be undertaken to better understand several issues with its use, such as investigating the human understanding of global versus local optimization predictions, limiting the scope of collaboration to avoid overwhelming operators in time critical situations, and implementing different predictive elements that would prevent or at least mitigate the fixation problems seen in this study.

- The super active level clearly had performance advantages over other levels of automation in this experiment. Therefore, it is desirable to try and leverage this high level of performance for routine tasks while employing a different level of decision support or some sort of active mitigation strategy to reduce the subject's errors of commission for critical events. Some type of adaptive automation could be employed that automatically reduces the level of re-planning when unexpected events occur.

- More investigation into should be undertaken into operator coping strategies. An interesting observation was that operators followed several well defined coping strategies to deal with cognitive overload, such as cognitive shedding. Research focused on understanding these failure modes is needed to determine if

any particular interface elements influence the coping strategy chosen by operators and how these could be used in a positive vein. This type of research could guide future design of interfaces to support operators even while in highly off-nominal conditions.

Bibliography

[1] David S. Alberts, John J. Garstka, and Frederick P. Stein. *Network Centric Warfare: Developing and Leveraging Information Superiority.* CCRP Publication Series. Command and Control Research Program (CCRP), Washington, DC, 2nd edition, 2000.

[2] H. H. Bell and D. R. Lyon, editors. *Using Observer Ratings to Assess Situation Awareness.* Lawrence Erlbaum Associates, Mahwah, NJ, 2000.

[3] C.E. Billings. *Aviation Automation: The Search for a Human-Centred Approach.* Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[4] A. Chapanis, F. C. Frick, W. R. Garner, J. W. Gebhard, W. F. Grether, R. H. Henneman, W. E. Kappaif, E. B. Newman, and A. C. Williams. Human Engineering for An Effective Air Navigation and Traffic Control System. National Research Council, Washington, D.C., 1951.

[5] M.L. Cummings. Automation Bias in Intelligent Time Critical Decision Support Systems. In *AIAA 1st Intelligent Systems Technical Conference Proceedings,* Chicago, IL, 2004. AIAA.

[6] M.L. Cummings. The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations. In *Adaptive Displays Conference Proceedings,* Los Angeles, CA, 2004.

[7] M.L. Cummings and S. Guerlain. An Interactive Decision Support Tool for Real-Time In-Flight Replanning of Autonomous Vehicles. In *AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop and Exhibit Proceedings*, Chicago, IL, 2004. AIAA.

[8] M.L. Cummings and S. Guerlain. Using A Chat Interface As An Embedded Secondary Tasking Tool. In *2nd Annual Human Performance, Situation Awareness, and Automation (HPSAA) Conference Proceedings*, Daytona Beach, FL, 2004.

[9] M.L. Cummings and S. Guerlain. Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles. *Presence*, In Review, 2005.

[10] DoD. Department of Defense Dictionary of Military and Associated Terms. Technical Report Joint Publication 1-02, 12 April 2001.

[11] DoD. Network Centric Warfare: Department of Defense Report to Congress. Technical report, Office of the Secretary of Defense, 21 July 2001.

[12] Mark H. Draper and Heath A. Ruff. Multi-Sensory Displays and the Visualization Techniques Supporting the Control of Unmanned Aerial Vehicles. In *IEEE International Conference of Robotics and Automation Proceedings*, San Francisco, CA, 2000.

[13] M.R. Endsley. Design and Evaluation for Situation Awareness Enhancement. In *Proceedings of the the Human Factors Society 32nd Annual Meeting*, volume 1, pages 97–101, Santa Monica, CA, 1988. Human Factors Society.

[14] M.R. Endsley. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1):32–64, 1995.

[15] Michael Freed, Robert Harris, and Michael G. Shafto. Human vs. Autonomous Control of UAV Surveillance. In *AIAA 1st Intelligent Systems Technical Conference Proceedings*, Chicago, IL, 2004. AIAA.

[16] V.J. Gawron. Human factors Issues in the Development, Evaluation, and Operation of Uninhabited Aerial Vehicles. In *Association for Unmanned Vehicle Systems International Conference (AUVSI '98) Proceedings*, pages 431–438, 1998.

[17] Michael Goodrich, Morgan Quigley, and Keryl Cosenzo. Task Switching and Multi-Robot Teams. In *Third International Multi-Robot Systems Workshop Proceedings*, Washington, DC, 2005.

[18] S. Guerlain. Interactive Advisory Systems. In *Proceedings of the Human Performance, Situation Awareness and Automation Conference*, pages 166–171, Savannah, GA, 2000.

[19] S. Hart and L. Staveland. Development of the NASA-TLX: Results of Empirical and Theoretical Research. In N. Meshkati, editor, *Human Mental Workload*, pages 139–183. North Holland, Amsterdam, 1988.

[20] D.B. Kaber, M.R. Endsley, and E. Onal. Design of automation for telerobots and the effect on performance, operator situation awareness and subjective workload. *Human Factors and Ergonomics in Manufacturing*, 10(4):409–430, 2000.

[21] C. Layton, P. J. Smith, and C. E. McCoy. Graphical Interfaces For Cooperative Planning Systems. In *Proceedings of the 1990 Conference on Systems, Man and Cybernetics*, pages 625–631, 1990.

[22] Michael Linegang, Craig Haimson, Jean MacMillan, and Jared Freeman. Human Control in Mixed-Initiative Systems: Lessons from the MICA-SHARC Program. In *8th International Conference on Systems, Man and Cybernetics Proceedings*, Arlington, VA, 2003. IEEE.

[23] Azad M. Madni and Carla C. Madni. Context-Driven Collaboration During Mobile C^2 Operations. In *The Society for Modeling and Simulation International 2004 Western Simulation Multiconference Proceedings*, San Diego, CA, 2004.

[24] Y. Miyata and D.A. Norman. Psychological Issues in Support of Multiple Activities. In D.A Norman and S.W. Draper, editors, *User Centered System Design:*

New Perspectives on Human Computer Interaction, pages 265–284. Lawrence Erlbaum, Hillsdale, NJ, 1986.

[25] M. Mouloua, R. Gilson, J. Kring, and P. Hancock. Workload, Situation Awareness, and Teaming Issues for UAV/UCAV Operations. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, pages 162–165, Minneapolis/St. Paul, MN, 2001. HFES.

[26] Jr. Olsen, Dan R. and Michael Goodrich. Metrics for Evaluating Human-Robot Interactions. In E.R. Messina and A.M. Meystel, editors, *PERMIS 2003 Proceedings*, Gaithersburg, MD, 2003. NIST.

[27] Jr. Olsen, Dan R. and Stephen Bart Wood. Fan-out: Measuring Human Control of Multiple Robots. In *Proceedings of the Conference on Human Factors in Computing Systems*, volume 6, pages 231–238, Vienna, Austria, 2004. ACM Press.

[28] G. Osga, K. Van Orden, N. Campbell, D. Kellmeyer, and D. Lulue. Design and Evaluation of Warfighter Task Support Methods in a Multi-Modal Watchstation. Technical Report 1874, SPAWAR Systems Center, 2002.

[29] Raja Parasuraman, Thomas B. Sheridan, and Chris D. Wickens. A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3):286–297, 2000.

[30] David L. Rockwell. Sensing the Future of UAVs. *AIAA Aerospace America*, September 2003.

[31] M. D. Rodgers, R. H. Mogford, and B. P. Strauch. Post Hoc Assessment of Situation Awareness in Air Traffic Control Incidents and Major Aircraft Accidents. In D. J. Garland, editor, *Situation Awareness Analysis and Measurement*, pages 73–102. Lawrence Erlbaum Associates, Mahwah, NJ, 2000.

[32] Emily M. Roth, Mark L. Hanson, Chris Hopkins, Vince Mancuso, and Greg L. Zacharias. Human in the Loop Evaluation of a Mixed-Initiative System for Planning and Control of Multiple UAV Teams. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, pages 280–284, Santa Monica, CA, 2004. Human Factors and Ergonomics Society.

[33] W. B. Rouse. *Systems Engineering Models of Human-Machine Interaction*. New York, 1983.

[34] Heath A. Ruff, G.L. Calhoun, M.H. Draper, J.V. Fontejon, and B.J. Guilfoos. Exploring Automation Issues in Supervisory Control of Multiple UAVs. In *2nd Human Performance, Situation Awareness, and Automation Conference (HP-SAA II) Proceedings*, Daytona Beach, FL, 2004.

[35] Heath A. Ruff, S. Narayanan, and Mark H. Draper. Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. *Presence*, 11(4):335–351, 2002.

[36] D. K. Schmidt. A Queuing Analysis of the Air Traffic Controller's Workload. *IEEE Transactions on Systems, Man, and Cybernetics*, 8(6):492–498, 1978.

[37] B. Schneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, MA, 1987.

[38] Thomas B. Sheridan. *Telerobotics, Automation and Human Supervisory Control*. The MIT Press, Cambridge, MA, 1992.

[39] Thomas B. Sheridan and W.L. Verplank. Human and Computer Control of Undersea Teleoperators. Man-Machine Systems Laboratory Report, MIT, 1978.

[40] P. Tsang and G. F. Wilson. Mental Workload. In G. Salvendy, editor, *Handbook of Human Factors and Ergonomics*, pages 417–489. John Wiley and Sons, Inc., New York, 2nd edition, 1997.

[41] Amos Tversky and Daniel Kahneman. Judgement under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[42] K. Vicente. *Cognitive Work Analysis. Toward Safe, Productive, and Healthy Computer-Based Work.* Lawrence Erlbaum Associates, Mahwah, New Jersey, 1999.

[43] C. D. Wickens. Situation Awareness: Impact of Automation and Display Technology. In *Situation Awareness: Limitations and Enhancements in the Aviation Environment*, Neuilly Sur Seine, France, 1995.

[44] Chris D. Wickens, Stephen Dixon, and Dervon Chang. Using Interference Models to Predict Performance in a Multiple-Task UAV Environment - 2 UAVs. Technical Report AHFD-03-9/MAAD-03-1, Aviation Human Factors Division, Institute of Aviation, University of Illinois at Urbana-Champaign, April 2003.

[45] Glenn F. Wilson and Christopher A. Russell. Psychophysiologically Determined Adaptive Aiding In A Simulated UCAV Task. In *Human Performance Situation Awareness and Automation Technology II Conference Proceedings*, Daytona Beach, FL, 2004.

[46] R. M. Yerkes and J. D. Dodson. The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, 18:459–482, 1908.

Appendix A

MAUVE Rules of Engagement (RoE)

The following are specific rules that governed how scenarios and UAV actions in MAUVE operated.

A.1 Naming Conventions

- UAVs

  - Numbered 1-4

- Targets

  - T-XXP where XX = target number and P = priority

  - Priority may be High (H), Medium (M), or Low (L)

  - Examples: T-1H, T-12M, T-23L

- Waypoints (WP)

  - WP-XY where X = UAV# the waypoint is associated with and Y = waypoint letter

  - Examples: WP-1A, WP-2C

- Loiterpoints (LP)

    - LP-XY where X = UAV# the loiterpoint is associated with and Y = loiterpoint letter

    - Examples: LP-4E, LP-3H

- Threats/Hazards

    - H-XX where XX = threat number

    - Example: H-1, H-12

A.2   Arm Payload (Arming)

- Arming must be initiated within an arming or firing window, and only when the UAV is at the desired target. The system also will not allow arming to be started if it will not finish by the end of the firing window.

- All arming windows are 10 seconds long and always immediately precede the beginning of a firing window.

- Arming takes 5 seconds +/- 2 seconds to complete (range 3-7 seconds). The value was chosen at random from this range.

A.3   Fire Payload (Firing)

- Firing must occur within a firing window (TOT window), the UAV must have previously been armed, and it only can happen when the UAV is located at the desired target. The system also will not allow firing if it will not finish by the end of the firing window.

- Arming may occur in the firing window, though this indicates that UAV is late and in danger of missing its deadline, but firing may not occur in the arming window.

- All firing windows are 20 seconds long and always immediately follow the corresponding arming window.

- Firing takes 5 seconds +/- 2 seconds to complete (range 3-7 seconds). The value was chosen at random from this range.

A.4 Battle Damage Assessment (BDA)

- BDA must be scheduled prior to destroying a target, and thus cannot be added or removed after the arming window for that particular target has started.

- If BDA is scheduled to occur, it will occur automatically after firing (no user interaction required). However, if firing does not occur, neither will BDA, regardless of whether is it scheduled or not.

- By default, all high priority targets require BDA and all medium/low priority targets do not, unless re-planning during the scenario causes changes to this convention.

- BDA takes 45 seconds to complete, and once started must be finished.

A.5 Miscellaneous

- If a UAV reaches an active target (where active is defined as not destroyed and TOT not passed), it will automatically loiter at the target. Otherwise, the UAV will continue to the next target without stopping.

- In general, when time is short, destroy higher priority targets instead of lower priority targets.

Appendix B

Supplemental Experiment Screens

The screens on the following page were seen by participants throughout the test scenarios on the two MMWS monitors not being used by the MAUVE simulation interface.

OBJECTIVES

1 Return to base (RTB) within the time limit for the mission.

2 Comply with changing mission requirements, which will be relayed to you by periodic intelligence messages, such as a RTB order earlier than the time limit.

3 Destroy all targets before their time on target (TOT) window ends.

4 Perform battle damage assessment (BDA) on specified targets after destroying them.

5 Avoid taking damage from enemies by navigating around and out of threat areas.

6 Answer communications.

Figure B-1: The Objectives Screen, Far Left MMWS Monitor

COLOR CODING

UAV Action	Color
Enroute	Blue
Loitering	Orange
Arming Payload	Yellow
Firing Payload	Red
Battle Damage Assessment	Brown
Return to Base	Green

Figure B-2: The Color Coding Table, Bottom Center MMWS Monitor

Appendix C

Descriptive Statistics

C.1 Demographics

Category	N	Min.	Max.	Mean	Std. Dev.
Age (years)	12	20	42	26.33	6.97
Military Experience (years)	9	1	20	5.67	6.38
UAV Experience (hours)	2	1	20	12.50	10.61
Pilot/Flight Experience (hours)	9	5	400	119.44	140.86
Student (Y/N)	10	-	-	-	-
Gender (M/F)	10M, 2F	-	-	-	-

Table C.1: Descriptive Statistics for Study Demographics

C.2 Dependent Measures

Dependent Measure	Min.	Max.	Mean	Std. Dev.
Performance Score	122	1000	734.053	234.275
Number of TOT Delay Requests	1	41	11.542	11.575
Prop. TOT Delay Requests Approved	0	1	0.541	0.354
Interaction Wait Time (WTI)	19	299	81.542	57.599
Wait Time in the Queue WTQ	0	182	35	43.364
Situation Awareness Wait Time (WTSA)	10	831	263.708	239.730
Total System Wait Time (WTT)	31	890	346.083	237.774
Situation Awareness	1	5	3.688	1.111
Erroneous Target Critical Events	0	2	0.333	0.565
Subjective Workload	38	84.667	62.625	12.897
Secondary Workload	21	58.67	36.740	10.368

Table C.2: Descriptive Statistics for all Dependent Measures

Appendix D

Performance Score

D.1 Overview

The performance score was designed to be an overall measure of session performance. It is independent of re-planning level, or how many events were planned to occur in a specific scenario. An individual's score was based upon the total number and type of mission objectives they completed over the course of an entire session (Appendix H), while penalties were assessed for actions that had negative consequences for the mission plan, such as a UAV taking hits or incorrectly destroying a target. Each completed mission objective or penalty event had a particular amount of points associated with it, negative or positive, from which the performance score was calculated as 1000 times the number of points accumulated, divided by the total number possible in that test session.

D.2 Earned Points

Earned points in the performance score were accumulated when the operator successfully completed various mission objectives. The base number of points earned in the performance score for achieving an objective (Table D.1) corresponded to how long

Event	Base Points
Target Correctly Destroyed	30
BDA Correctly Performed	45

Table D.1: Base Points Earned For Completion of Primary Mission Objectives

finishing that objective took, or was scheduled to take in the scenario. For example, performing BDA took 45 seconds, so it was worth 45 base points, while destroying a target was worth 30 base points because that was the length of the arming plus firing windows for each target in the test sessions. This was then modified by the priority of the target itself and the difficulty of that target (Table D.2). Before taking into account target difficulty, medium priority targets were set to be worth 1.5 times a low priority target (45 points) and high priority targets twice as much (60 points). These ratios were consistent with the rules of engagement and prioritized objectives given to subjects in the simulation. The specific numbers were chosen because the average target priority throughout all of the test scenarios was medium, which corresponded to an average of 45 points per target destroyed, the same as BDA. As completing both types of actions awarded the same average number of points and targets took a maximum of 30 seconds to destroy while BDA took 45 seconds to complete, this meant that subjects' time was more valuable spent destroying targets.

Throughout the test scenarios. there was significant variability in how difficult it was to destroy a target. Factors influencing how difficult a target was to destroy included whether it was affected by re-planning, the number of simultaneous events occurring near to the target's TOT, and whether it was at the start, middle or end of a route. To account for this in the performance score, each target in the test sessions was assigned a difficulty of low, medium, or high based upon a combination of two independent, expert observer ratings and experiment data on how often each target was missed by operators. This data is presented in Table D.3 for the low re-planning scenario, and Table D.4 for the high re-planning scenario. To avoid biasing

Priority	Difficulty	Modified Points
	Hard	67.5
High	Medium	60
	Low	52.5
	Hard	52.5
Medium	Medium	45
	Low	37.5
	Hard	37.5
Low	Medium	30
	Low	22.5

Table D.2: Modified Points Earned for Target Destruction

the performance scores between different scenarios, 1/3 of the targets were assigned each to each difficulty level, and if the targets could not be divided equally between the difficulty ratings, an additional number of medium difficulty ratings for targets were assigned.

The base point differential between targets of differing priorities was 15 per level of difference (i.e. 15 unless the two targets being compared were low and high priorities, in which case the difference was 2 levels, or 30 points). As it was determined that priority was a more important indicator of a target's relative value to the mission plan than difficulty, the number of points the difficulty adjustment could modify a target's score up or down was half of this difference (7.5 points). Therefore, if a target was assigned a difficulty rating of easy, it was worth 7.5 points less than normal, a medium difficulty rating resulted in no change, and a hard rating increased its value by 7.5 points. For a complete listing of target values based on priority and difficulty rating, see Table D.2. This modification was small enough that a lower priority target could never be worth more than a higher priority target, or vice versa. For example, an easy difficulty high priority target was worth 52.5 points, while a hard difficulty medium priority target was worth the same, as can be seen in Table D.2.

BDA points were also modified based upon difficulty, because even though BDA took a long time to perform, in most cases it was already scheduled at mission launch

Target Name	Incompletion Rate	Observer 1 Rating	Observer 2 Rating	Combined Rating
T-1M	0	L	H	M
T-2H	0	L	L	L
T-3H	0	L	M	L
T-4L	0	M	L	M
T-5H	0	L	L	L
T-6H	0	L	M	M
T-7M	0	L	L	L
T-8L	0	M	H	H
T-9M	0.08	H	M	H
T-10M	0.08	H	H	H
T-11L	0.08	H	M	H
T-12L	0	M	L	L
T-13L	0	M	M	M
T-14H	0.08	H	M	M
T-15H	0.75	H	H	H
T-16M	0	L	L	L
T-17L	0.08	H	M	M
T-18M	0	L	H	M
T-19M	0	M	H	H

Table D.3: Target Difficulty Ratings for the Low Re-planning Scenario

and did not require any attention on the part of the operator. This made it a very easy task for operators to perform. The only challenging part of BDA for operators was accommodating changes in BDA into their schedule at the last minute, as was the case for re-planning events. Therefore, BDA events were either classified as being difficult or easy based upon whether the event was affected by re-planning or not. If a difficult BDA event was completed, it was worth the full 45 base points, but completion of easier, automated BDA events was modified to award only half points (22.5).

D.3 Penalty Points

Penalty points were assessed when the operator performed actions or guided at least one of the UAVs under their supervision into a state that ran counter to mission

Target Name	Incompletion Rate	Observer 1 Rating	Observer 2 Rating	Combined Rating
T-1L	0.25	H	H	H
T-2M	0	L	L	L
T-3H	0.17	L	M	L
T-4H	0.50	H	M	H
T-5H	0.33	M	L	M
T-6M	0.08	M	L	L
T-7H	0.25	M	M	M
T-8L	0.17	M	M	M
T-9L	0.58	H	H	H
T-10H	0.42	M	H	H
T-11M	0.25	L	M	M
T-12L	0.17	L	L	L
T-13L	0.17	L	M	M
T-14L	0.33	H	L	M
T-15M	0.17	M	L	L
T-16H	0	M	L	L
T-17M	0.25	H	M	H
T-18M	0.08	L	H	M
T-19M	0.83	H	H	H

Table D.4: Target Difficulty Ratings for the High Re-planning Scenario

objectives. The four types of possible events that incurred penalty points were 1) destroying a target when there previously had been a command not to do so, 2) incorrectly performing BDA, 3) having a UAV enter a threat area, or 4) arriving at base beyond the mission time limit. The penalties assessed, just like for points earned, were based upon the time it took to accomplish that particular event, and are listed in Table D.5. As destroying a target due to missed or ignored orders is a critical event in command and control situations, operators were doubly penalized for doing this in the performance score. Subjects did not receive points for destroying inappropriate targets, and were penalized 45 additional points each time this error occurred. This number of penalty points was chosen because it was equal to what an average target was worth in the scenario, and in these types of situations, the priority of the target itself becomes the order not to fire the weapon. The same double penalty did not

Event	Penalty Points
Target Incorrectly Destroyed	45
BDA Incorrectly Performed	0
Threat Hit	10
Late Arrival to Base	1 per second

Table D.5: Penalty Points for Actions Counter to Mission Objectives

apply when the operator incorrectly performed BDA on a target. In that case, the operator did not receive points for taking the action, as it was of no use to the mission, and the wasted time spent performing that BDA would translate into time penalties in subsequent events. In following with the one second per point reasoning outlined above, late arrivals to base were penalized at a rate of 1 point per second, per vehicle, and hits on any UAV at 10 per hit, because threats fired on UAVs in threat areas every 10 seconds.

Appendix E

Subjective Situation Awareness (SA) Score

E.1   Indicators of Situation Awareness

The subjective SA score was based upon what were judged to be the four most salient indicators of situation awareness in the test sessions. These indicators were:

1. The number of entries into threat areas where the UAV received 3 or more hits and the operator did not intervene to minimize further damage to that UAV.

2. The amount of system wait time at targets due to loss of situation awareness.

3. The number of targets missed due to lack of situation awareness.

4. The percentage of re-planning events successfully completed.

The first two indicators measured level 2 SA (comprehension), as a UAV loitering unnecessarily at a target or a UAV traveling through a threat area could be perceived directly from the map display. The second two indicators measured level 3 SA (future projection), as a failure to successfully complete a re-plan or a missing a target due

to a lack of SA meant that the operator had not properly executed, modified and projected the mission plan in response to scenario events.

E.2   Rating Scales and Scoring

The range of possible values for each of the SA indicators listed above were grouped and ranked on a 1-5 scale (Table E.1). The scale range was chosen so the subjective SA score could easily be compared to the objective SA score developed for this study (responses to the online questions). The groupings were based on subjective expert observations of all the experimental trials. Each test session was assigned four individual scores, one for each indicator, and the average of these ratings was the subjective SA score for that test session. A score for Level 2 SA was obtained by averaging the ratings for indicators 1 and 2, while a score for Level 3 SA was obtained similarly with indicators 3 and 4.

As an example of how to calculate the subjective SA scores for a trial, consider the following data from one of the test sessions: 0 entries into threat areas, 95 seconds of system wait time at targets due to loss of SA, 3 targets missed due to lack of SA, and 86% of all re-plans completed successfully. From table E.1, this would result in ratings of 5, 3, 4, and 4 for the SA indicators listed above, giving an overall subjective SA score of (5+3+4+4)/4 = 4, a subjective level 2 SA score of (5+3)/2 = 4, and a

Subjective Situation Awareness Score	Number entries into threat areas, no operator intervention	Amount of time UAVs spent loitering at missed or removed targets (s)	Number of targets missed due to lack of SA	Percentage of re-plans successfully completed
5	0	0	0-1	90 or more
4	-	0-30	2-3	80-90
3	1	30-120	4-5	70-80
2	-	120-300	6-7	60-70
1	2 or more	300 or more	8 or more	60 or less

Table E.1: Subjective SA Indicator Rating Scales

subjective level 3 SA score of $(4+4)/2 = 4$ for this test session.

Appendix F

COUHES Consent Form

CONSENT TO PARTICIPATE IN
NON-BIOMEDICAL RESEARCH

Management of Multiple Dynamic Human Supervisory Control Tasks

You are asked to participate in a research study conducted by Professor Mary Cummings Ph.D, from the Aeronautics and Astronautics Department at the Massachusetts Institute of Technology (M.I.T.). You were selected as a possible participant in this study because the expected population this research will influence is expected to contain men and women between the ages of 18 and 50 with an interest in using computers. You should read the information below, and ask questions about anything you do not understand, before deciding whether or not to participate.

PARTICIPATION AND WITHDRAWAL

Your participation in this study is completely voluntary and you are free to choose whether to be in it or not. If you choose to be in this study, you may subsequently withdraw from it at any time without penalty or consequences of any kind. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

PURPOSE OF THE STUDY

The study is designed to evaluate how increasing levels of automated scheduling support affect an operator's performance and situation awareness while supervising multiple simultaneous dynamic tasks, and how this changes as workload increases. Situation awareness is generally defined as the perception of the elements in the environment, the comprehension of the current situation, and the projection of future status of the related system.

PROCEDURES

If you volunteer to participate in this study, we would ask you to do the following things:

- Attend a training and practice session to learn a video game-like software program that will have you supervising and interacting with multiple unmanned aerial vehicles.

- Practice on the program will be performed until an adequate level of performance is achieved, which will be determined by your demonstrating basic proficiency in monitoring the vehicles, redirecting them as necessary, executing commands such as firing and arming of payload at appropriate times, and responding to online instant messages (estimated time 1.5 hours).

- Execute two thirty minute trials consisting of the same tasks as above (1 hour)

- Attend a debriefing to determine your subjective responses and opinion of the software (10 minutes).

- All testing will take place in MIT building 37, rooms 301 and 307.

- Total time: 2-3 hours, depending on skill level.

POTENTIAL RISKS AND DISCOMFORTS

There are no anticipated physical or psychological risks in this study.

POTENTIAL BENEFITS

While there is no immediate foreseeable benefit to you as a participant in this study, your efforts will provide critical insight into the human cognitive capabilities and limitations for people who are expected to supervise multiple complex tasks at once, and how system automation can help overcome these barriers.

PAYMENT FOR PARTICIPATION

You will be paid $10/hr to participate in this study which will be paid upon completion of your debrief. Should you elect to withdraw in the middle of the study, you will be compensated for the hours you spent in the study.

CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law. You will be assigned a subject number which will be used on all related documents to include databases, summaries of results, etc. Only one master list of subject names and numbers will exist that will remain only in the custody of Professor Cummings.

IDENTIFICATION OF INVESTIGATORS

If you have any questions or concerns about the research, please feel free to contact the Principal Investigator, Mary L. Cummings, at (617)-252-1512, e-mail, missyc@mit.edu, and her address is 77 Massachusetts Avenue, Room 33-305, Cambridge, MA 02139. The student investigator is Paul J. Mitchell and he may be contacted by telephone at (617)-452-3043 or via email at pmitchel@mit.edu.

EMERGENCY CARE AND COMPENSATION FOR INJURY

"In the unlikely event of physical injury resulting from participation in this research you may receive medical treatment from the M.I.T. Medical Department, including emergency treatment and follow-up care as needed. Your insurance carrier may be billed for the cost of such treatment. M.I.T. does not provide any other form of compensation for injury. Moreover, in either providing or making such medical care available it does not imply the injury is the fault of the investigator. Further infor-

mation may be obtained by calling the MIT Insurance and Legal Affairs Office at 1-617-253-2822."

RIGHTS OF RESEARCH SUBJECTS

You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E32-335, 77 Massachusetts Ave, Cambridge, MA 02139, phone 1-617-253-6787.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

_____

Name of Subject

_____   _____

Signature of Subject   Date

SIGNATURE OF INVESTIGATOR

In my judgment the subject is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

_____   _____

Signature of Investigator   Date

Appendix G

Demographic Survey

MAUVE Demographic Survey

1. Age: _____

2. Gender: ☐ Male ☐ Female

3. Occupation: _____

   If student:

   (a) *Class Standing:* ☐ Undergraduate ☐ Graduate

   (b) *Major:* _____

   If currently or formerly part of any country's armed forces:

   (a) *Country/State:* _____

   (b) *Status:* ☐ Active Duty ☐ Reserve ☐ Retired

   (c) *Service:* ☐ Army ☐ Navy ☐ Air Force ☐ Other: _____

   (d) *Rank:* _____

   (e) *Years of Service:* _____

4. Do you have experience with remotely piloted vehicles (land, sea, air)?

   ☐ Yes
   ☐ No

   If yes:

   (a) *Type/class:* _____

   (b) *Number of hours:* _____

5. Do you have piloting experience other than with remotely piloted vehicles?

☐ Yes
☐ No

If yes:

(a) *Type/class:* _____

(b) *Number of hours:* _____

6. How often do you play video games?

☐ Never
☐ Less than 1 hour per week
☐ Between 1 and 4 hours per week
☐ Between 1 and 2 hours per day
☐ More than 2 hours per day

7. Are you color blind?

☐ Yes
☐ No

Appendix H

MAUVE Instructions

Introduction

Thank you in advance for your participation! Today you will participate in an experiment designed to evaluate how effectively an operator can control multiple unmanned aerial vehicles (UAVs), given increasing levels of automation assistance. This session will be divided into 3 phases:

1. *Training*

   You'll be put through 3-4 interactive training scenarios designed to illustrate possible situations you may encounter during testing. At any time during training, please feel free to ask any and all questions you may have about the interface and how to interact with it. This will last approximately 45-90 minutes, depending upon how quickly you learn.

2. *Testing*

   You will then participate in two 30-minute trials similar to those seen during training. This will take approximately 70 minutes.

3. *Post-Test Feedback*

   Your feedback on the interface will be solicited through a focused interview and post-evaluation survey. This will take a maximum of 15 minutes.

Background

In this experiment, you are an unmanned aerial vehicle (UAV) operator that is responsible for supervising 4 UAVs that are collectively tasked with destroying a set of time-sensitive targets in a suppression of enemy air defenses mission. The area contains enemy threats capable of firing on your UAVs. The UAVs are highly autonomous, and therefore only require high level mission planning and execution from you. The UAVs launch with a pre-determined mission plan that comes from an air tasking order (ATO), so initial target assignments and routes have already been completed for you. Your job will be to monitor their progress, re-plan aspects of the mission in reaction to unexpected events, and in some cases manually execute mission critical actions such as arming and firing of payloads.

UAVs are capable of 6 high-level types of actions in the simulation: traveling enroute to targets, loitering at specific locations, arming payloads, firing payloads, performing battle damage assessment, and returning to base. Battle damage assessment (otherwise known as battle damage imagery or BDI) is the post-firing phase of combat where it is determined whether the weapon(s) hit the target, and if the desired effect was achieved. Table 1 outlines the color coding assigned to each of these actions in the simulation.

UAV Action	Color
Enroute	Blue
Loitering	Orange
Arming Payload	Yellow
Firing Payload	Red
Battle Damage Assessment	Brown
Return to Base	Green

Table 1: Color coding of UAV actions in MAUVE

You have two main objectives in this experiment: 1) To guide each UAV's actions so that together, all UAVs under your supervision properly execute the required missions and engagements of the up-to-date ATO, which will change over time, and 2) To answer periodic questions about the situation from commanders.

Supervising the UAVs to meet ATO specifications is your primary task and must be handled first, so don't answer questions when any of the UAVs require your attention. Supervision of the UAVs can be broken down into the following prioritized sub-tasks, from highest priority to lowest:

1. Return to base (RTB) within the time limit for the mission (this limit will be clearly marked).

2. Comply with changing mission requirements, which will be relayed to you by periodic intelligence messages, such as a RTB order earlier than the time limit.

3. Destroy all targets before their time on target (TOT) window ends.

4. Perform battle damage assessment (BDA) on specified targets after destroying them.

5. Avoid taking damage from enemies by navigating around and out of threat areas.

6. Answer communications.

These sets of objectives will often conflict with one another. In these cases, you must perform the actions that have the highest priority first.

Example Communications

Below are some sample questions that will give you an idea of the kind of knowledge that will be required of you during scenarios.

- How many medium priority targets remain to be destroyed?

- Which UAVs will arrive at their next checkpoint by 12:05:30?

- How many targets have TOTs ending in the next 7 minutes?

- Which threat area is adding the most time to a single UAV's mission plan?

- With the current mission plan, how many active targets will not be destroyed?

Rules of Engagement (RoE)

The following are specific rules that govern how scenarios and UAV actions in MAUVE operate. You should keep this reference handy throughout your training until you are familiar with everything in this section.

Naming Conventions

- UAVs

  - Numbered 1-4

- Targets

  - T-XXP where XX = target number and P = priority
  - Priority may be High (H), Medium (M), or Low (L)
  - Examples: T-1H, T-12M, T-23L

- Waypoints (WP)

  - WP-XY where X = the UAV the waypoint is associated with and Y = waypoint letter
  - Examples: WP-1A, WP-2C

- Loiter Points (LP)

  - LP-XY where X = the UAV the loiter point is associated with and Y = loiter point letter
  - Examples: LP-4E, LP-3H

- Threats/Hazards

  - H-XX where XX = threat number
  - Example: H-1, H-12

Arm Payload (Arming)

- Arming must be initiated within an arming or firing window, and only when the UAV is at the desired target. The system also will not allow you to arm if it will not finish by the end of the firing window.

- All arming windows are 10 seconds long and always immediately precede the beginning of a firing window.

- Arming takes 5 seconds +/- 2 seconds to complete (range 3-7 seconds).

Fire Payload (Firing)

- Firing must occur within a firing window (TOT window), the UAV must have previously been armed, and it only can happen when the UAV is located at the desired target. The system also will not allow you to fire if it will not finish by the end of the firing window.

- Note that you may arm in the firing window (though this indicates you are late and in danger of missing your deadline), but you may not fire in the arming window.

- All firing windows are 20 seconds long and always immediately follow the corresponding arming window.

- Firing takes 5 seconds +/- 2 seconds to complete (range 3-7 seconds).

Battle Damage Assessment (BDA)

- BDA must be scheduled prior to destroying a target, and thus cannot be added or removed after the arming window for that particular target has begun.

- If BDA is scheduled to occur, it will occur automatically after firing (no user interaction required). However, if firing does not occur, neither will BDA, regardless of whether is it scheduled or not.

- By default, all high priority targets require BDA and all medium/low priority targets do not, unless re-planning during the scenario causes changes to this convention.

- BDA takes 45 seconds to complete, and once started must be finished.

Miscellaneous

- If a UAV reaches an active target (where active is defined as not destroyed and TOT not passed), it will automatically loiter at the target. Otherwise, the UAV will continue to the next target without stopping.

- In general, if you find yourself out of time, destroy higher priority targets instead of lower priority targets.

Appendix I

Example Trial Time Series

The following series of screenshots are an example of what subjects saw on their displays over the course of an entire test session, taken at approximately 2 minute intervals. The trial shown is with the super active automation level and the high re-planning scenario.
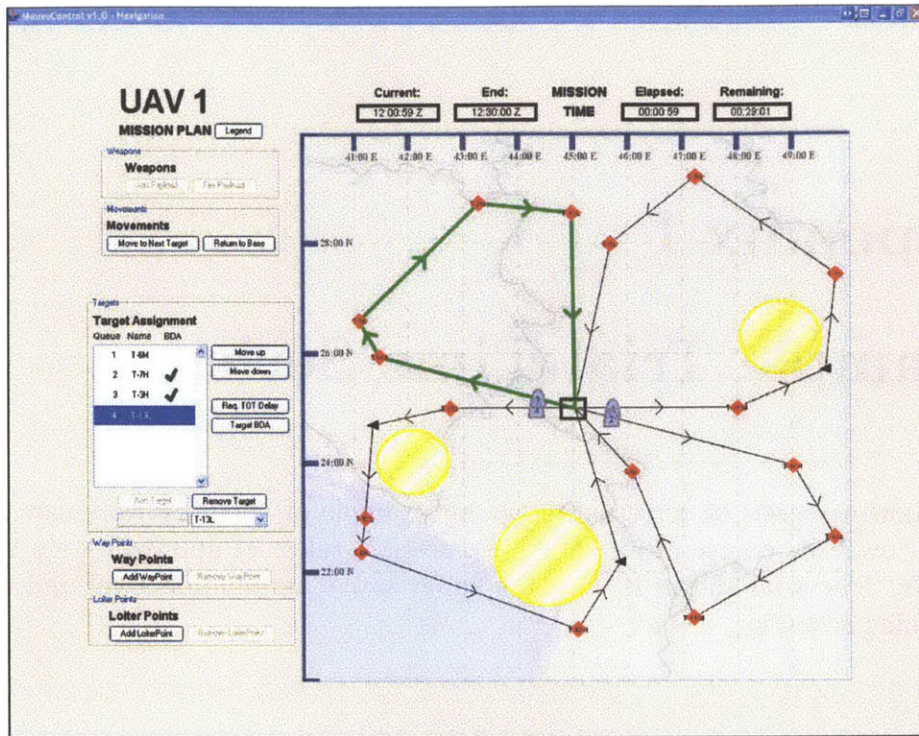
(a) Navigation Window
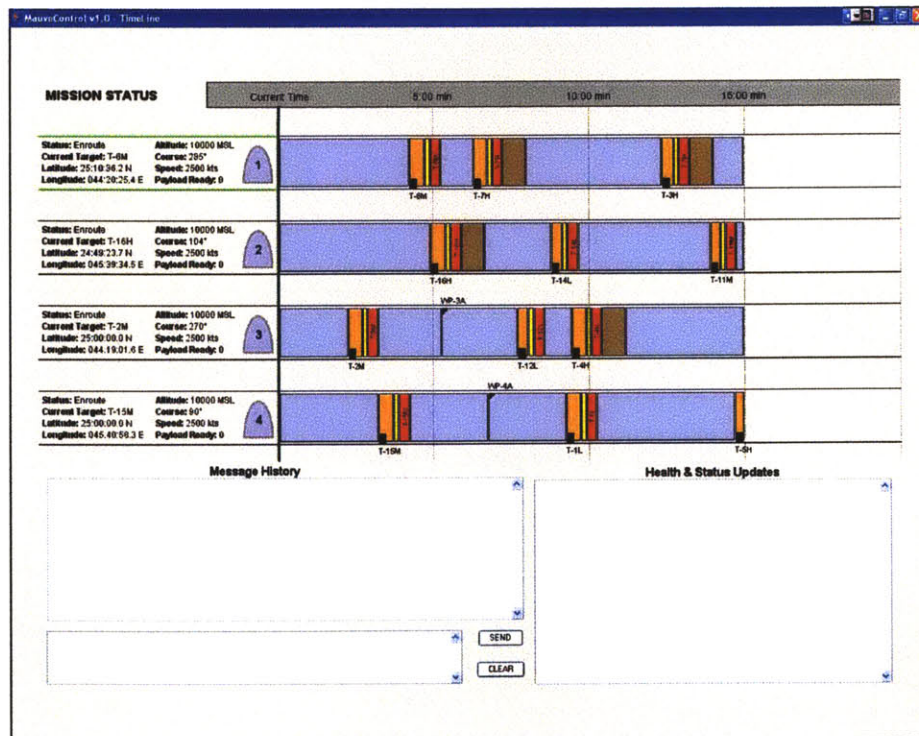
(b) Decision Support Window

Figure I-1: Super Active and High Re-planning Example - 12:00:59
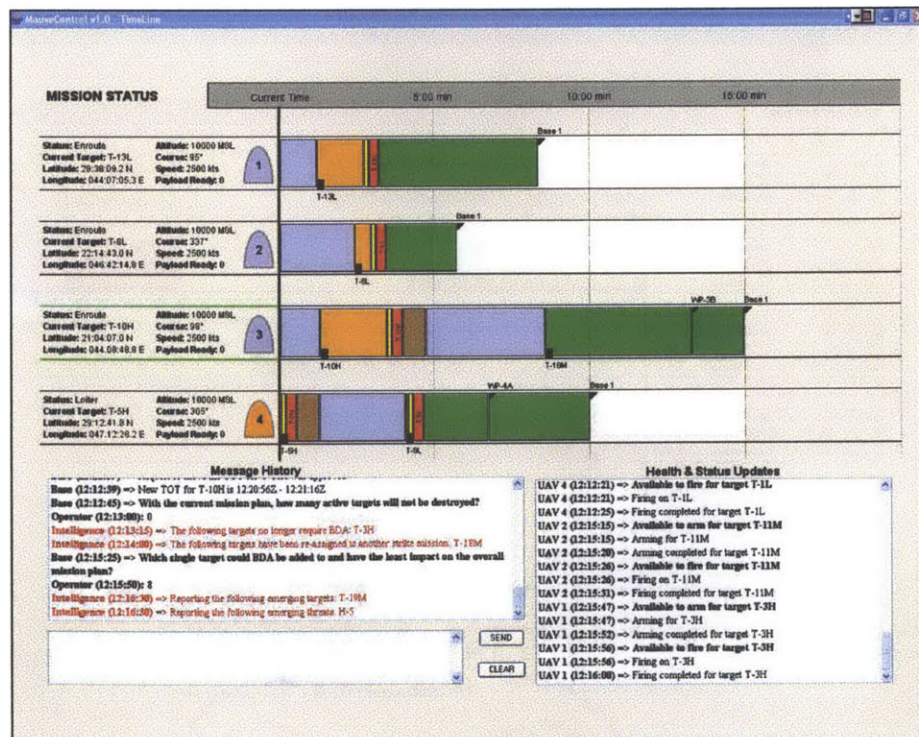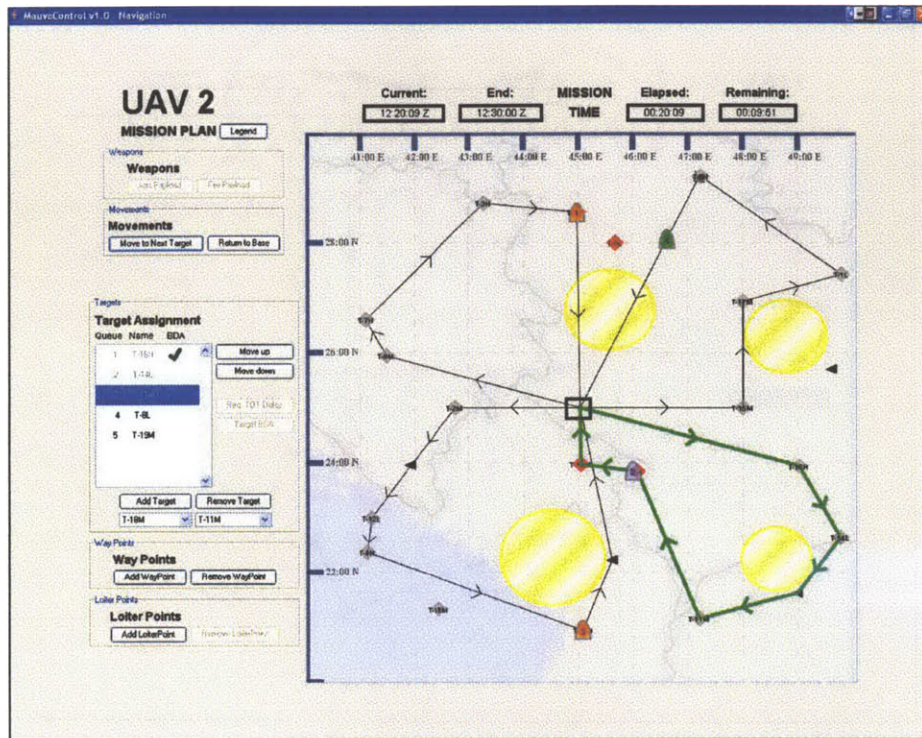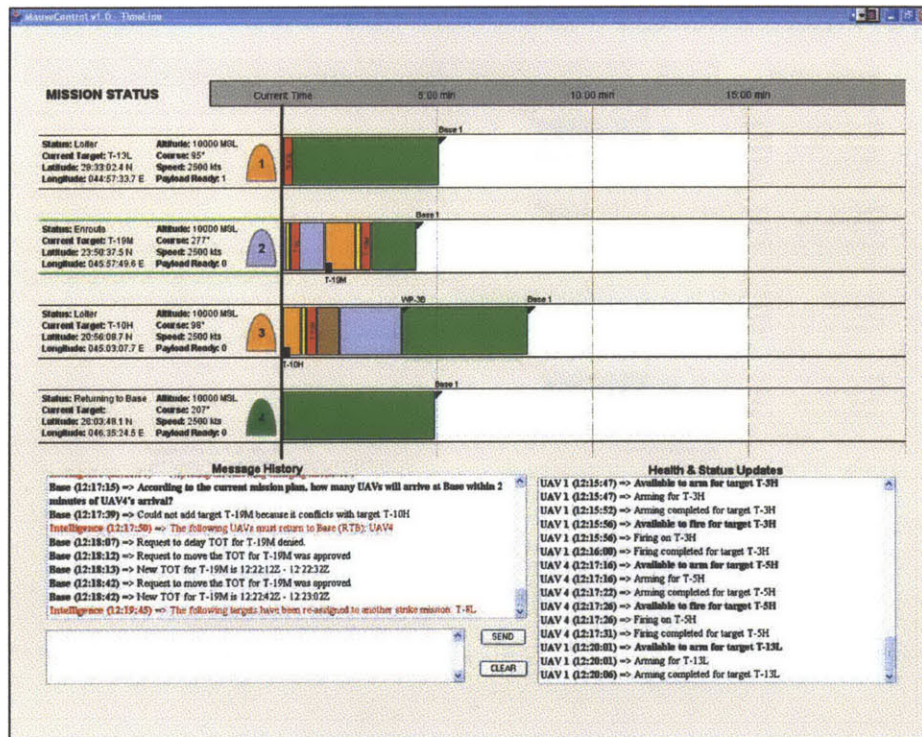
170
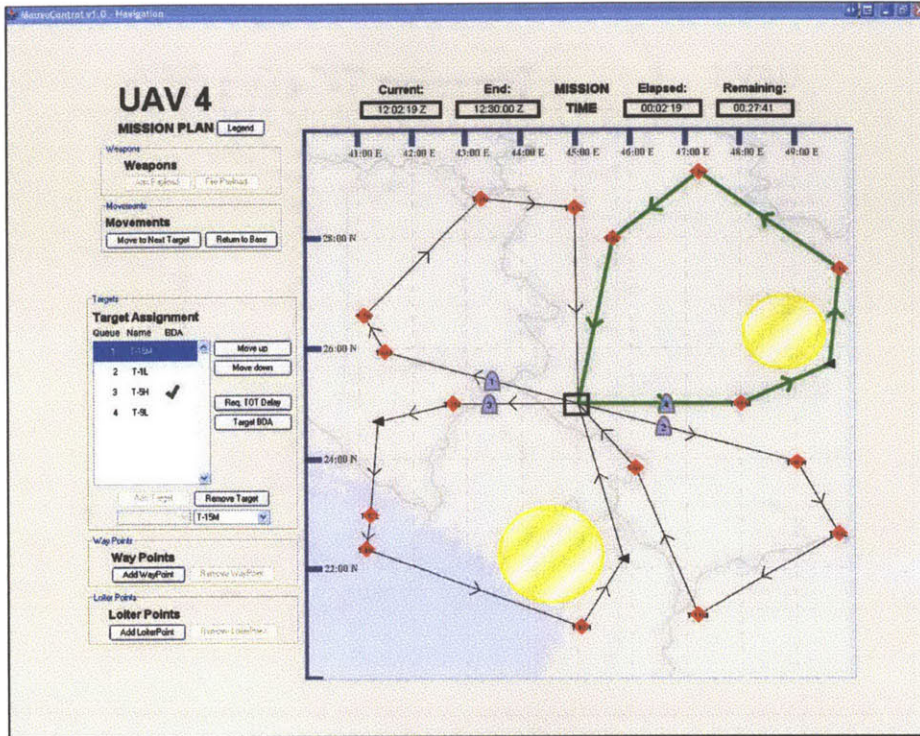
(a) Navigation Window
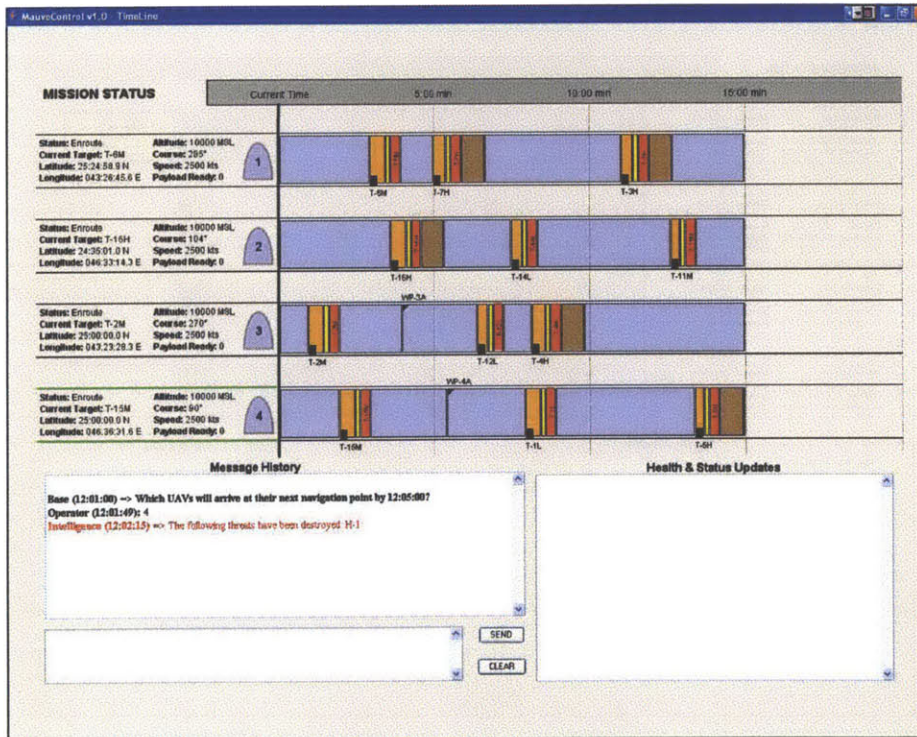
(b) Decision Support Window

Figure I-2: Super Active and High Re-planning Example - 12:02:19
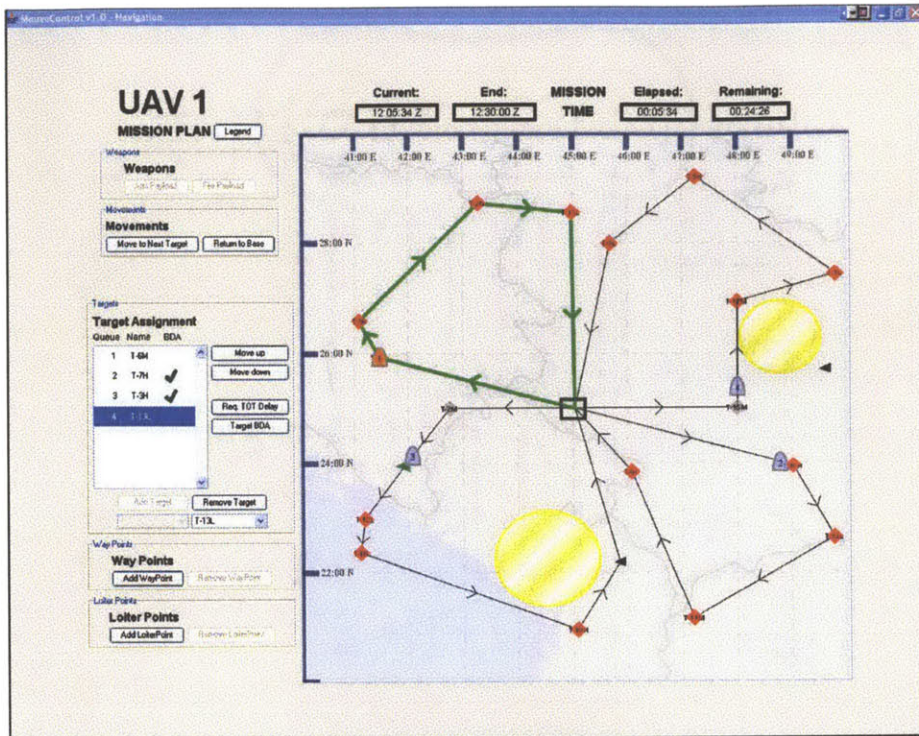
(a) Navigation Window
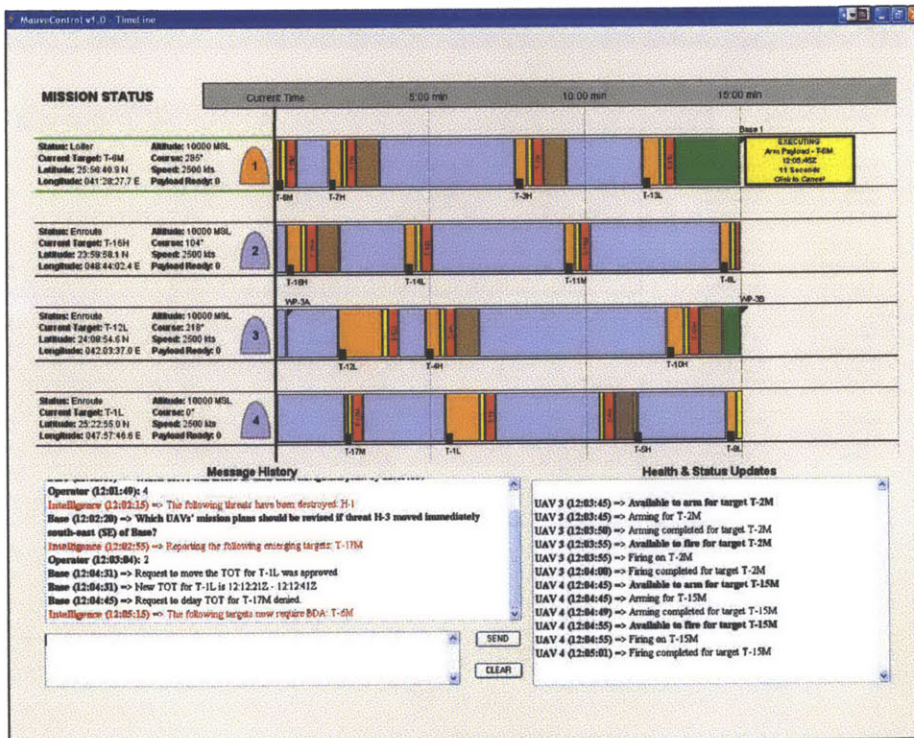
(b) Decision Support Window

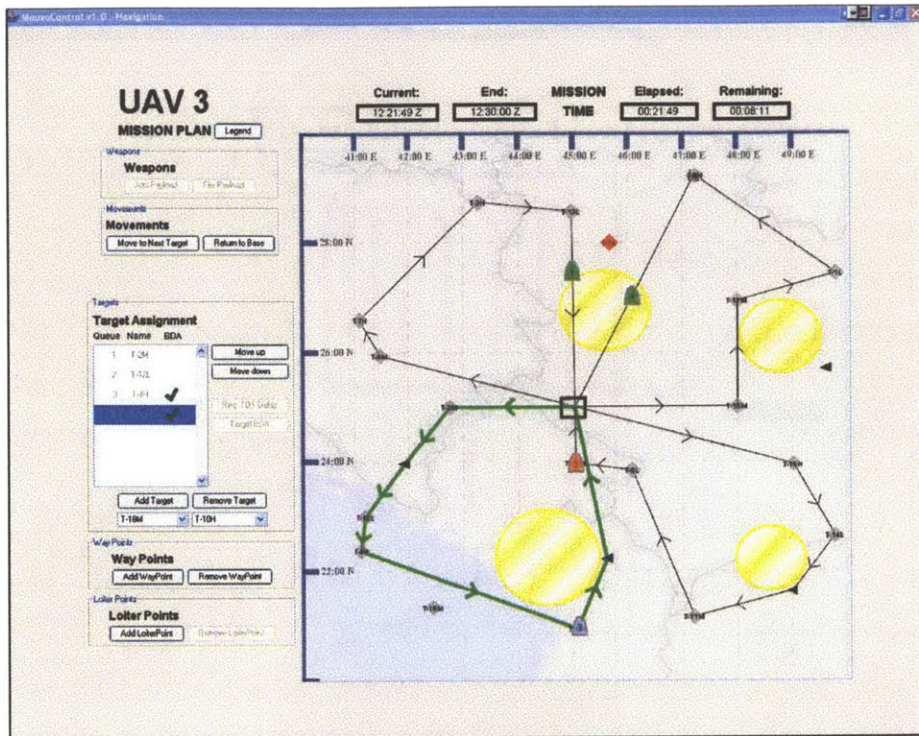Figure I-3: Super Active and High Re-planning Example - 12:05:34

172
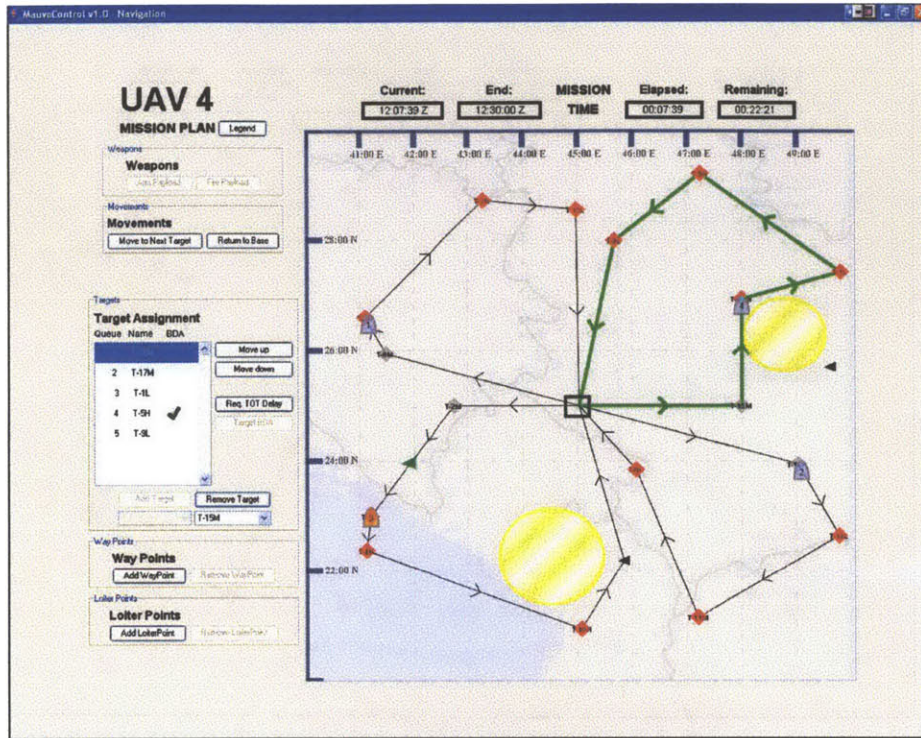
(a) Navigation Window
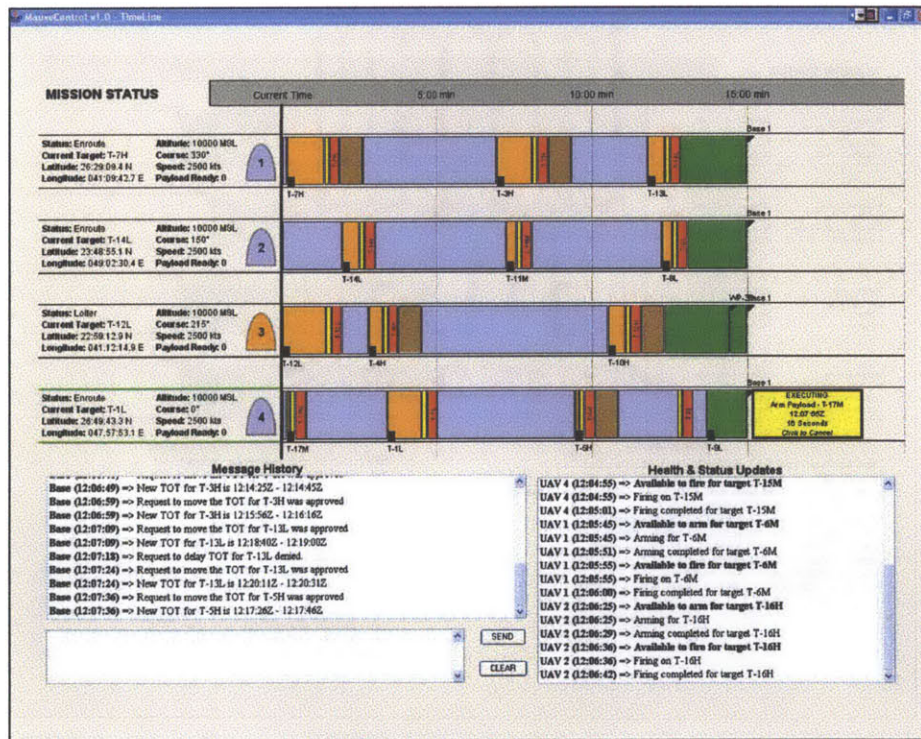
(b) Decision Support Window

Figure I-4: Super Active and High Re-planning Example - 12:07:39
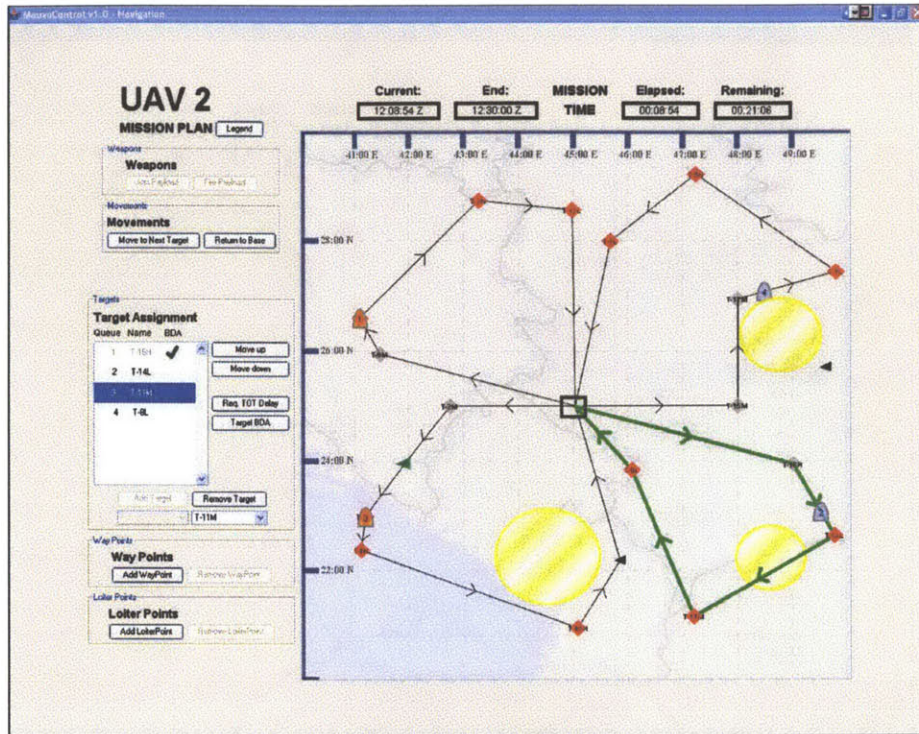
(a) Navigation Window
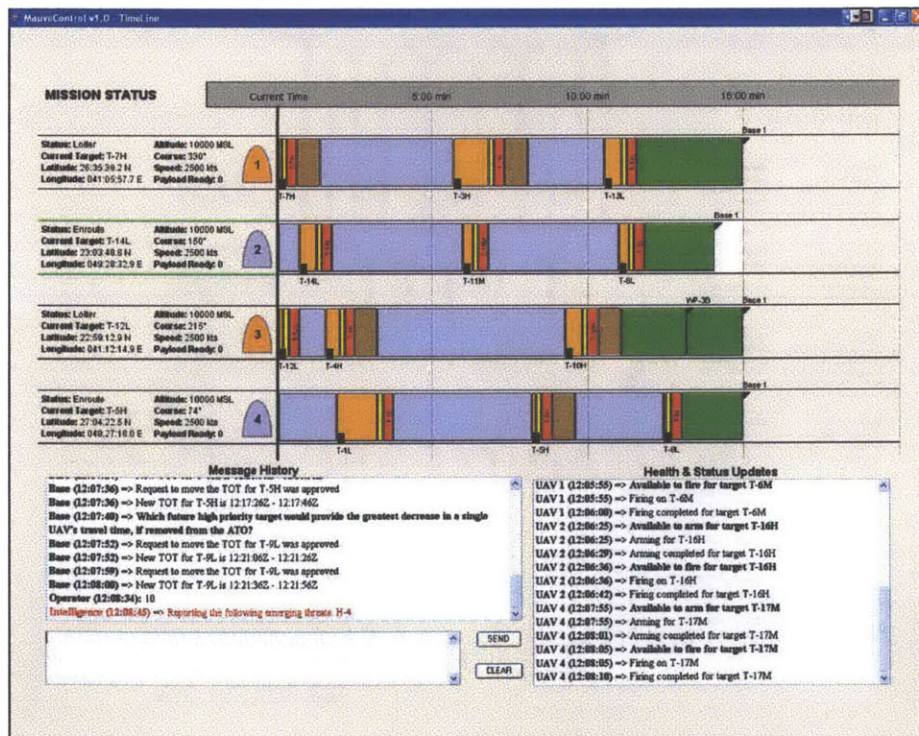
(b) Decision Support Window

Figure I-5: Super Active and High Re-planning Example - 12:08:54
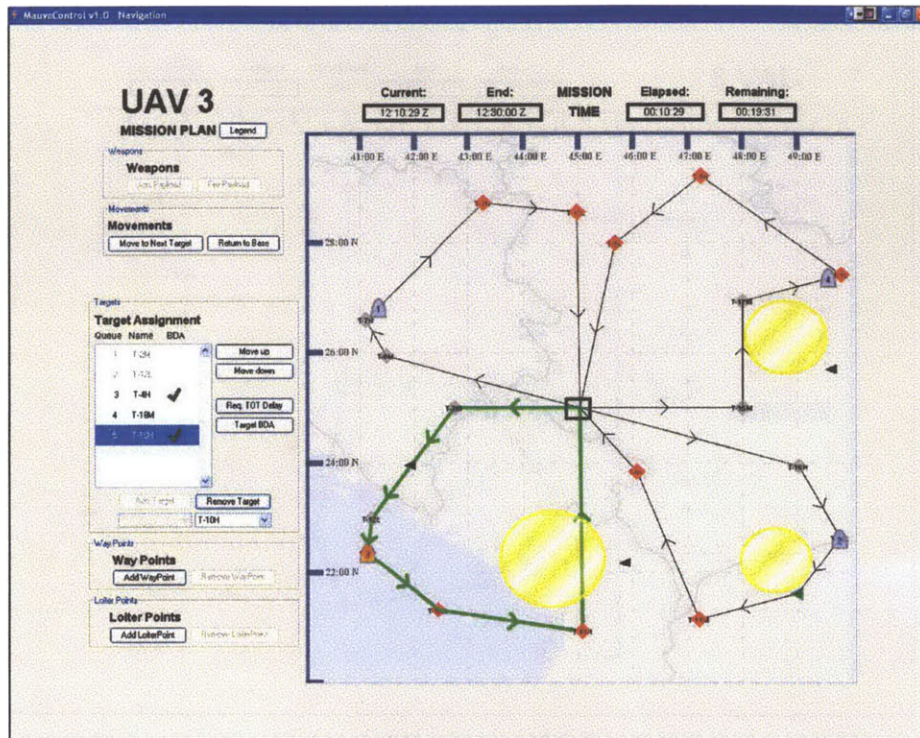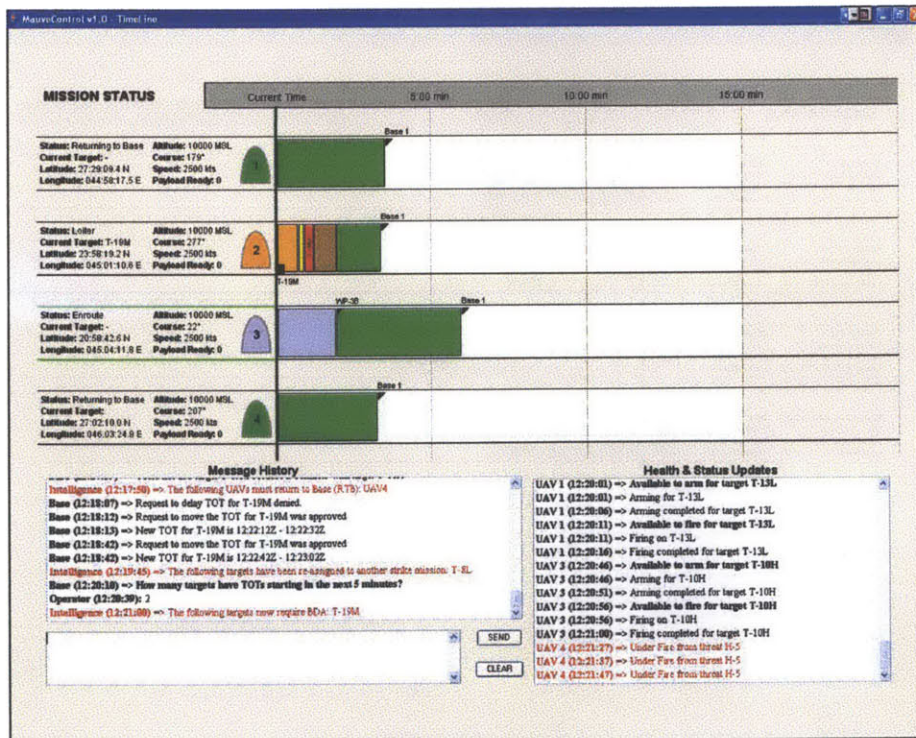
(a) Navigation Window
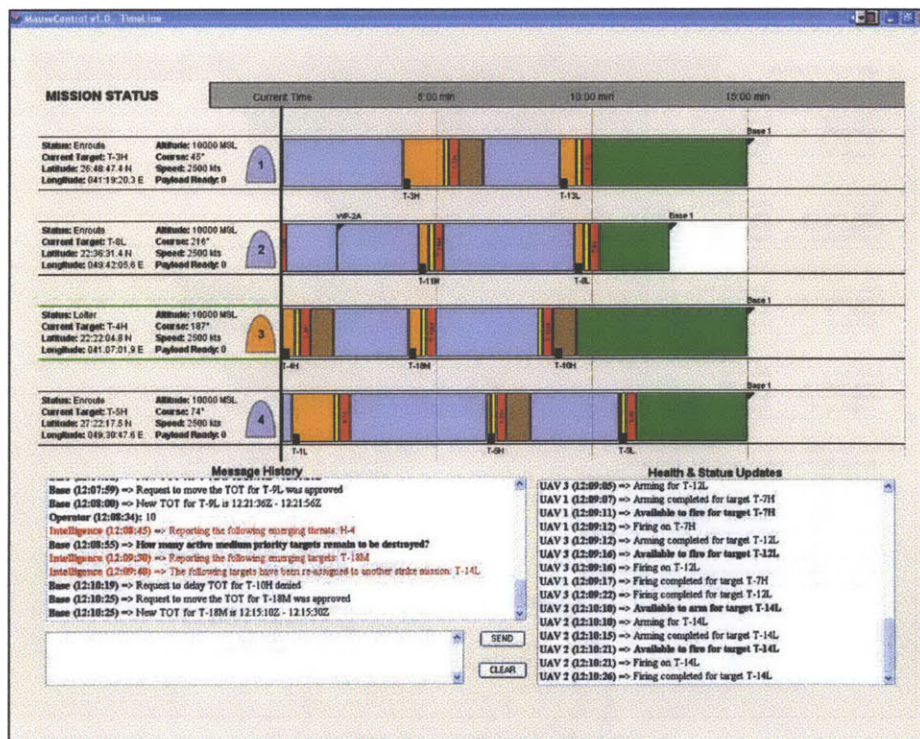
(b) Decision Support Window

Figure I-6: Super Active and High Re-planning Example - 12:10:29
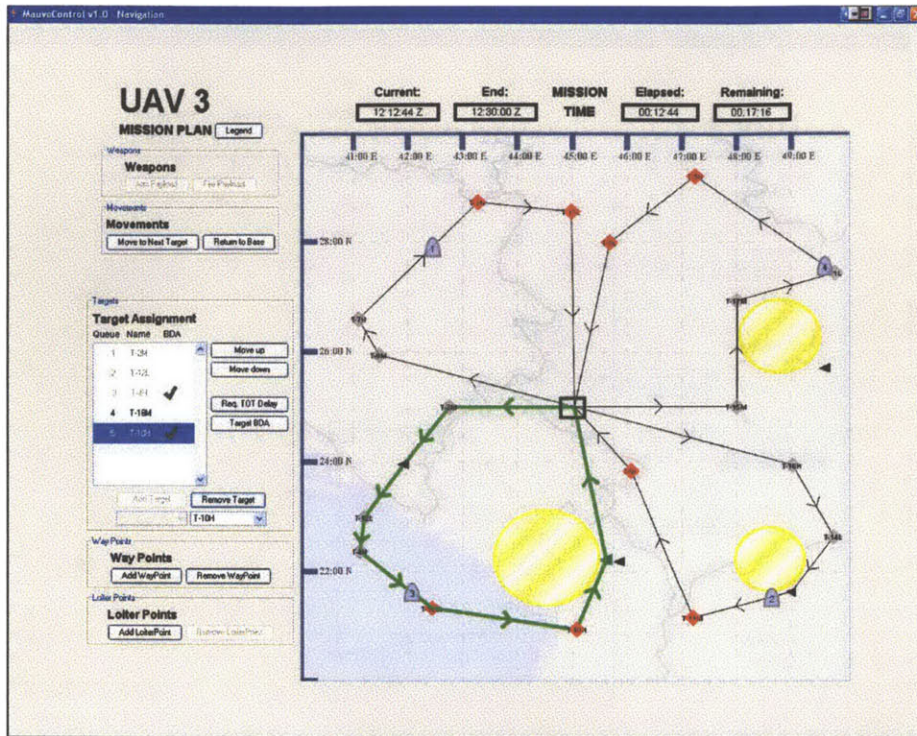
(a) Navigation Window
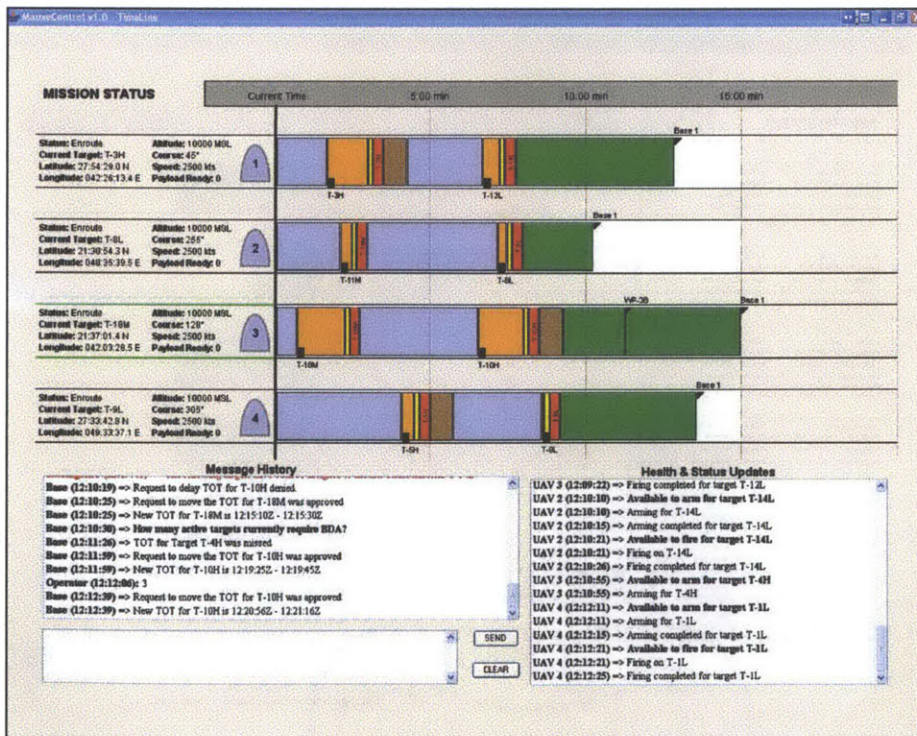
(b) Decision Support Window

Figure I-7: Super Active and High Re-planning Example - 12:12:44
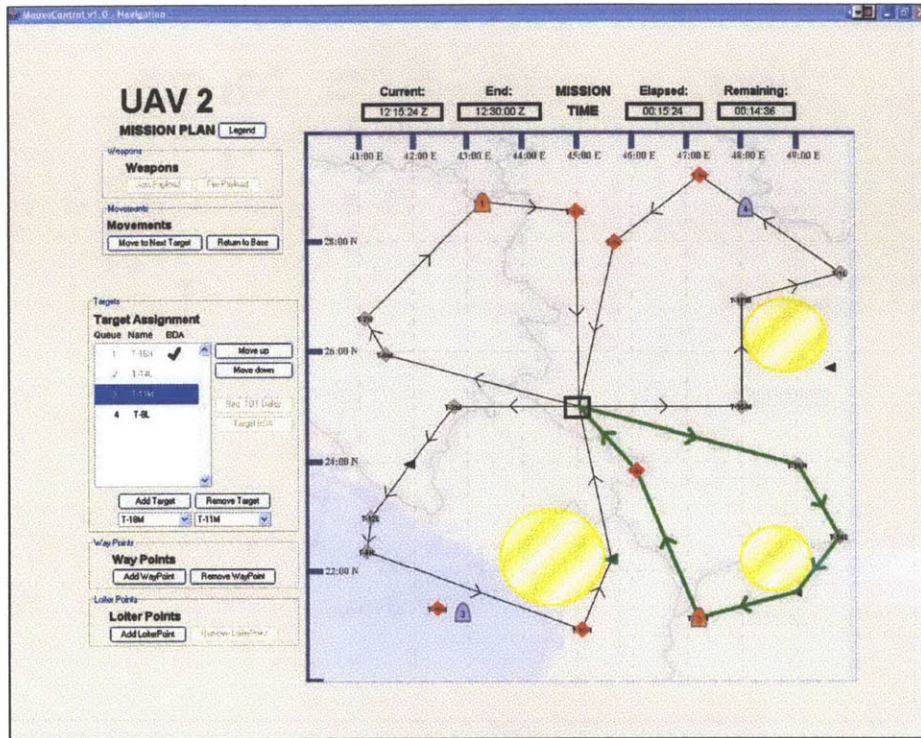
(a) Navigation Window
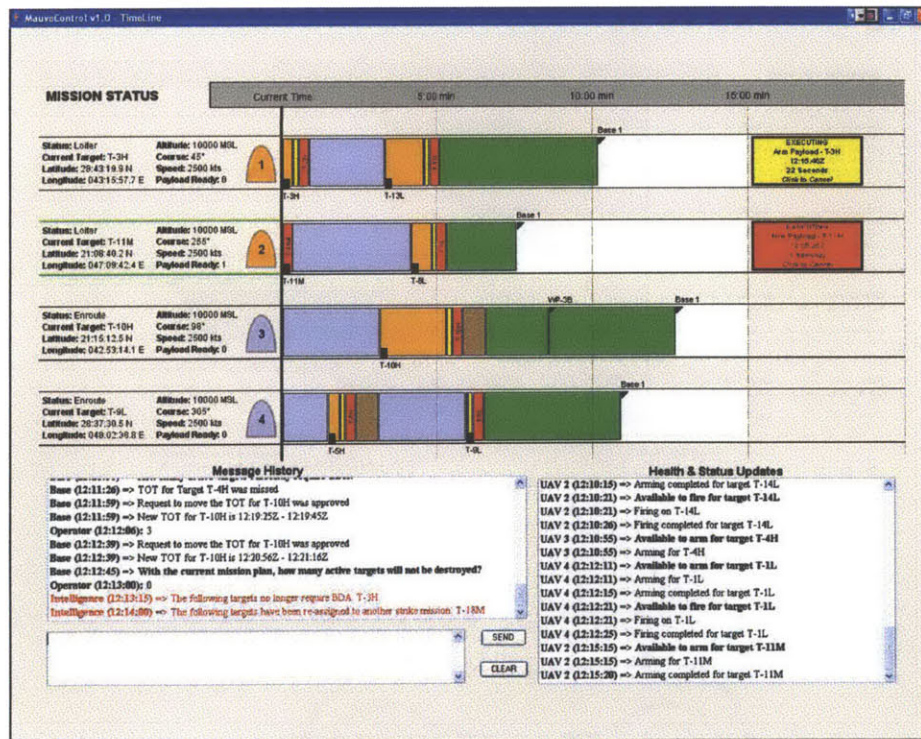
(b) Decision Support Window

Figure I-8: Super Active and High Re-planning Example - 12:15:24
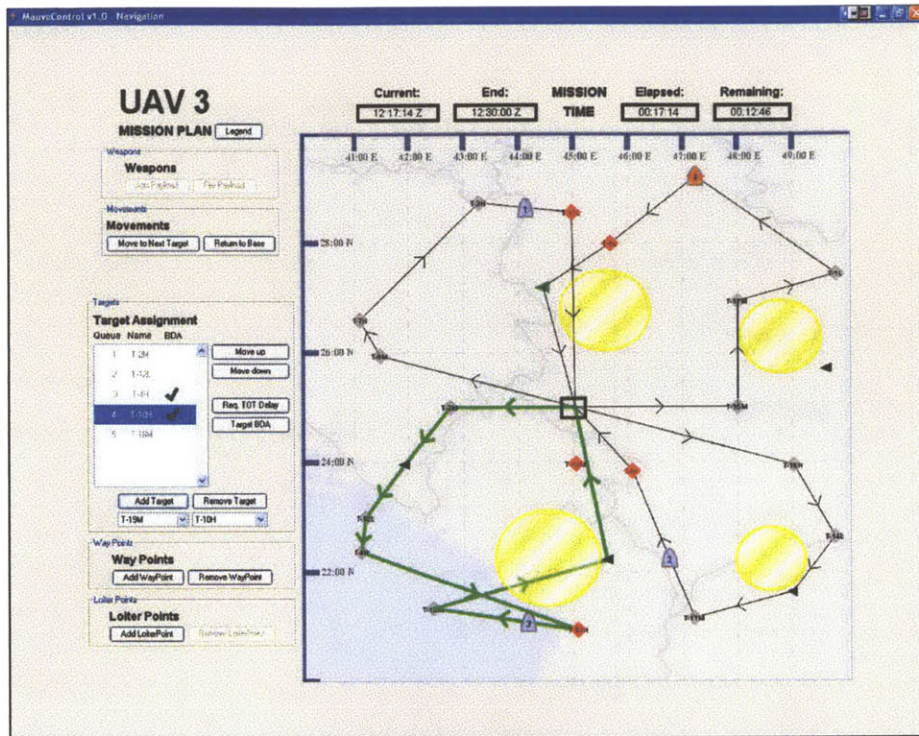
(a) Navigation Window

(b) Decision Support Window

Figure I-9: Super Active and High Re-planning Example - 12:17:14

(a) Navigation Window

(b) Decision Support Window

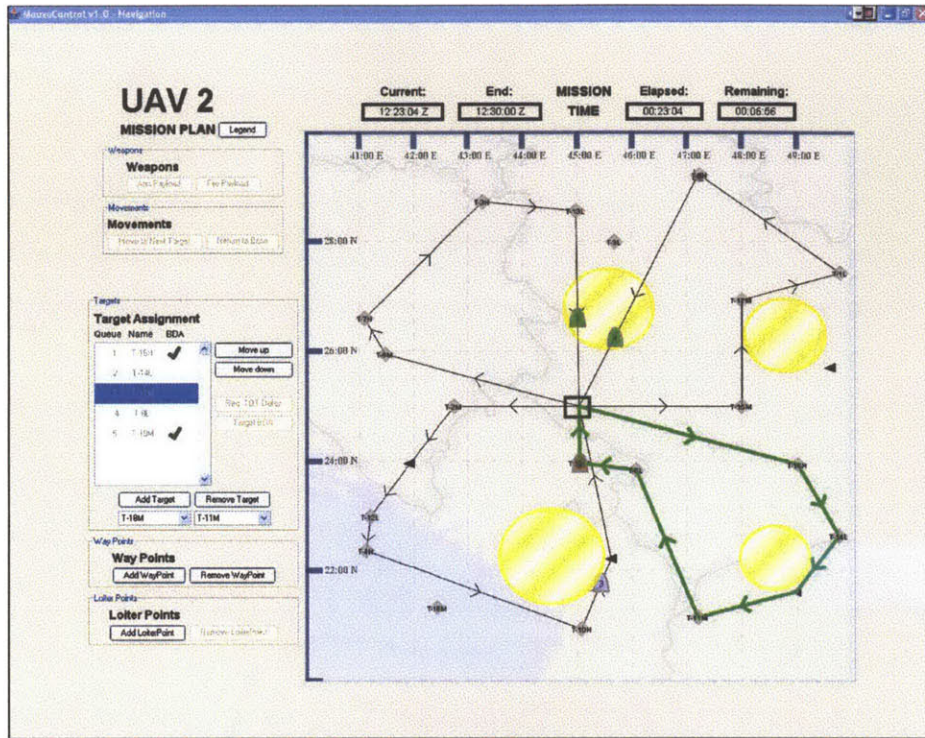Figure I-10: Super Active and High Re-planning Example - 12:20:09
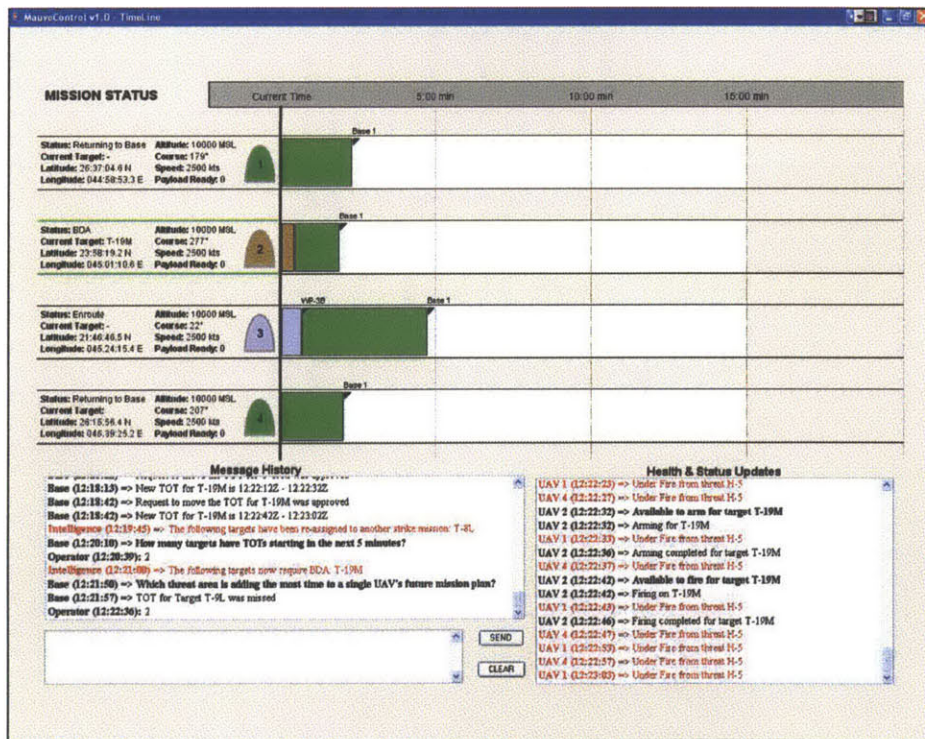
(a) Navigation Window

(b) Decision Support Window

Figure I-11: Super Active and High Re-planning Example - 12:21:49

(a) Navigation Window

(b) Decision Support Window

Figure I-12: Super Active and High Re-planning Example - 12:23:04