StarVis: A Configural Decision Support Tool for Schedule Management of Multiple Unmanned Aerial Vehicles

by

Amy S. Brzezinski

S.B. Aeronautical & Astronautical Engineering
Massachusetts Institute of Technology, Cambridge, MA, 2005

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

February 2008

Signature of Author: ______________________________
Department of Aeronautics and Astronautics
December 18, 2007

Certified by: ______________________________
Mary Cummings
Assistant Professor of Aeronautics and Astronautics
Thesis Supervisor

Accepted by: ______________________________
David L. Darmofal
Associate Professor of Aeronautics and Astronautics
Chair, Committee on Graduate Students

This Page Left Intentionally Blank

StarVis: A Configural Decision Support Tool for Schedule Management of Multiple Unmanned Aerial Vehicles

by

Amy S. Brzezinski

Submitted to the Department of Aeronautics and Astronautics
on December 18, 2007 in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Aeronautics and Astronautics

Abstract

As unmanned aerial vehicles (UAVs) become increasingly autonomous, current single-UAV operations involving multiple personnel could transition to a single operator simultaneously supervising multiple UAVs in high-level control tasks. These time-critical, single-operator systems will require advance prediction and mitigation of schedule problems to ensure mission success. However, actions taken to address current schedule problems may create more severe future problems. Decision support could help multi-UAV operators evaluate different schedule management options in real-time and understand the consequences of their decisions. This thesis describes two schedule management decision support tools (DSTs) for single-operator supervisory control of four UAVs performing a time-critical targeting mission. A configural display common to both DSTs, called StarVis, graphically highlights schedule problems during the mission, and provides projections of potential new problems based upon different mission management actions. This configural display was implemented into a multi-UAV mission simulation as two different StarVis DST designs, Local and Q-Global. In making schedule management decisions, Local StarVis displayed the consequences of potential options for a single decision, while the Q-Global design showed the combined effects of multiple decisions. An experiment tested the two StarVis DSTs against a no DST control in a multi-UAV mission supervision task. Subjects using the Local StarVis performed better with higher situation awareness and no significant increase in workload over the other two DST conditions. The disparity in performance between the two StarVis designs is likely explained by the Q-Global StarVis projective "what if" mode overloading its subjects with information. This research highlights how decision support designs applied at different abstraction levels can produce different performance results.

Thesis Supervisor: Mary Cummings
Title: Assistant Professor of Aeronautics and Astronautics

Acknowledgements

Many people deserve thanks for making my graduate experience at MIT so wonderful:

My advisor, Prof. Missy Cummings, who took a chance on a NASA graduate co-op. Your guidance, patience, and alliance throughout my research have been a highlight of my time at MIT. I aspire to be a mentor like you.

The MITRE Corporation, for funding the work represented in this thesis.

Stacy Scott, for inspiring my experimental question in a fifteen minute conversation.

Jake Crandall and Jill Drury, for providing feedback on this thesis.

The Desperados, Sylvain Bruni and Jessica Marquez, for helping a sometimes clueless Masters student survive graduate school.

Eric Jones, for making me laugh during the toughest times, inspiring me with his enthusiasm and work ethic, and making Green Dinosaur a nice hat.

Liang Sim, for inspiring excellence in all things, particularly in this thesis.

My officemates: Hudson Graham, Anna Massie, and Geoff Carrigan, for supporting me through the challenging time of thesis writing, and making every day seem like a party.

HAL labmates, post-docs, and visiting individuals not already mentioned: Dhiman Bhattacharjee, Yves Boussemart, Scott Fisher, Angela Ho, Brian Mekdeci, Carl Nehme, Patricia Pina, Cristin Smith, Chris Tsonis, Jordan Wan, Prof. Gilles Coppin, Prof. John Lee, Mark Ashdown, Enlie Wang, Fernanda Muzzio Almirao, Fernanda Borques da Silva, Carina Furusho, Julien Nicolas, Mauro Della Penna – thanks for your support through advice, suggestions, and happy hour trips to the R&D pub.

My UROPs, Albert Leung and Amanda Seybold, for doing all my dirty coding work and transforming my visualizations into reality.

Sian Kleindienst and Tamer Elkholy, for making Ashdown home.

Kathleen Connolly, for actually volunteering to keep me company during thesis writing.

Jack Batten, my high school physics teacher, who believed that I could thrive at MIT.

Erik Dill, for reminding me that the wonderful world beyond the walls of MIT is just waiting to be explored. You have changed not only my time at MIT but also who I am and how I see the world. I'm looking forward to our future adventures together.

And finally, my Mother, for supporting and encouraging me throughout my seven (wow) years at MIT. You inspire me more than you know.

Table of Contents

List of Figures

List of Tables

Nomenclature

ANOVA	Analysis of Variance
ATO	Air Tasking Order
BarVis	Bar Visualization
BDA	Battle Damage Assessment
BDI	Battle Damage Imagery
DoD	Department of Defense
DST	Decision Support Tool
MAUVE	Multi-Aerial Unmanned Vehicle Experiment test bed
MMWS	Multi-Modal Workstation
Multi-UAV	Multiple UAV
NASA TLX	National Aeronautics and Space Administration Task Load Index
OIF	Operation IRAQI FREEDOM
Q-Global	Quasi-Global
Req	Request
RoE	Rules of Engagement
RPV	Remotely Piloted Vehicle
RTB	Return to Base
SA	Situation Awareness
StarVis	Star Visualization
TOT	Time on Target
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
WOO	Window of Opportunity

1. Introduction

1.1. Motivation

An unmanned aerial vehicle (UAV) is defined as ”a powered, aerial vehicle that does not carry a human operator, uses aerodynamic forces to provide lift, can fly autonomously or be piloted remotely, can be expendable or recoverable, and can carry a lethal or non lethal payload” [1]. UAVs are increasingly used in a variety of military and civilian applications, such as reconnaissance, payload delivery, communication, surveillance, search and rescue, border patrol, and others [2]. UAVs have the advantage of being able to operate in contaminated or unsafe environments which could otherwise jeopardize the safety and lives of manned aircraft occupants. Additionally, UAVs can fly safely at lower or higher altitudes than manned aircraft [3], a capability useful for certain missions.

The United States has increased investment in military unmanned aerial system (UAS) development, procurement, and usage since the 1990s. From 1990 to 1999, the U.S. Department of Defense (DoD) invested over $3 billion in UAS development, procurement, and operations [4]. DoD increased UAS expenditure to $4.5 billion across the four year period from 2000 to 2004 [4]. In 2005, it was projected that spending on UAS would increase to a total of $3.1 billion in fiscal year 2011 [4]. UAV systems have been utilized in recent military operations such as Operation IRAQI FREEDOM (OIF) [5]. Throughout OIF, the Predator UAV system performed multiple types of missions such as intelligence, surveillance, reconnaissance, target designation, and strike [5]. Budgetary increases and the increased usage of UAVs in Iraq and Afghanistan indicate that UAV systems are becoming an important asset to the U.S. military.

Although physically unmanned, current UAVs are often controlled by teams of human operators who perform flying, navigation, and higher-level mission and planning tasks. However, as UAV flight and navigation tasks become more automated, UAV missions will likely transition from teams of people operating one UAV to one person supervising multiple UAVs. A single operator may be able to supervise and divide attention across multiple UAVs because automation will reduce the number of tasks requiring direct human control. The single operator's role will be one of supervisory control in which he or she will be responsible for high-level mission management tasks such as monitoring mission timelines and reacting to emergent events. This transition is advantageous for military operations, as it leads to a decrease in personnel.

Missions, especially those of a military nature, are often time-critical and tightly scheduled. In future time-critical, multi-UAV missions, the single operator's mental workload amount is of concern. Mental workload is a function of attention demand from numerous tasks. Periods of excessive workload could arise for an operator when critical tasks for several UAVs occur simultaneously, or if an operator needs to quickly and accurately switch between different tasks. Also of concern will be the effect of workload and increased automation on an operator's situation awareness. While increased automation is necessary to facilitate one operator to supervise multiple UAVs, it can increase mental workload and decrease situation awareness due to opacity, lack of feedback, and mode confusion [6] [7].

One of the primary tasks of a single operator supervising multiple UAVs in a time-critical mission will be to manage the mission schedule. In performing this task, it will be particularly important for operators to minimize future periods of excessive workload that could arise when tasks requiring operator action occur simultaneously. To a certain degree, it is possible to predict and mitigate high workload periods in advance. However, actions that eliminate a particular period of near-term high workload may create more severe high workload periods in the future, threatening mission success. Operators could have difficulty understanding the consequences of their schedule management decisions in the face of uncertainty, especially in dynamic environments in which they are performing other mission tasks. Thus, decision support is needed to help multi-UAV operators understand 1) the potential problems with their current mission schedule and 2) the consequences, both beneficial and detrimental, of actions taken to address those issues.

1.2. Problem Statement

The primary questions for this research effort are:

- Does a decision support tool that indicates 1) current schedule problems and 2) the potential consequences of fixing those problems help multi-UAV operators supervise the mission and effectively manage its schedule?

- What sort of scope or level of detail should decision support employ to provide multi-UAV schedule management information to operators?

- Does inclusion of schedule management decision support positively or negatively affect operator workload and situation awareness in multi-UAV supervision?

1.3. Research Objectives

In order to address the problem statement, the primary objective of this research is to develop and test a decision support tool (DST) for schedule management of multiple UAVs. This goal will be addressed through the following research objectives:

- Objective 1. **Develop a configural display for managing the schedules of multiple UAVs.** Configural displays, which are defined in Chapter 2, were chosen because they support efficient perceptual processes, which is necessary for decision-making under time pressure. In order to achieve this objective, a causal loop diagram was constructed to understand what variables are involved in multi-UAV schedule management decisions. Based upon this analysis, a configural display representing the variables by its form, called StarVis, was designed. Details on the causal loop diagram and the StarVis configural display are given in Chapter 3.

- Objective 2. **Implement the configural display into different decision support tool (DST) designs and embed the designs into a multi-UAV mission simulation.** The StarVis configural display was implemented in the Multi-Aerial Unmanned Vehicle Experiment (MAUVE) simulation test bed in two different DST designs: Local and Quasi-Global (Q-Global). The two StarVis implementations differed in how they represented the possible consequences of schedule management decisions. StarVis DST implementations are discussed in Chapter 3.

- Objective 3. **Evaluate the effect of the different configural DST implementations on human performance, workload, and situation awareness in a time-critical, multi-UAV supervision mission.** To address this objective, human subject experiments were performed with the Local and Q-Global StarVis configural DSTs embedded into the MAUVE simulation. Chapter 4 describes the experiment, while Chapter 5 presents the results and discusses how they answer the questions posed in the problem statement of this chapter.

1.4. Thesis Organization

The remainder of this thesis is organized as follows:

- Chapter 2, *Background,* summarizes past and current autonomy and human factors work in multiple UAV supervision with focus on scheduling challenges. It also discusses configural displays and frames the context of the research objectives introduced in Chapter 1.

- Chapter 3, *Star Visualization (StarVis) Configural Decision Support Tool,* presents the details of the StarVis configural display which was designed from influence diagram analysis of the variables involved in multi-UAV schedule management. This chapter also describes the implementation of the StarVis configural display into two DST designs, Local StarVis and Q-Global StarVis.

- Chapter 4, *Methods,* frames the experimental question and discusses experimental objectives, subjects, test bed, and procedures used in the MAUVE-StarVis human subject experiment.

- Chapter 5, *Results and Discussion,* presents the statistical analysis of the experiment described in the *Methods* chapter and discusses how the results answer the primary research questions of this study.

- Chapter 6, *Retrospective Analysis,* describes, presents, and discusses additional analysis performed to further explain and interpret the experimental results.

- Chapter 7, *Conclusion,* summarizes the motivation, objectives, and key findings of this research. Suggestions for future work are also provided.

2. Background

Recent multi-UAV research has either focused on algorithm development for completely autonomous multi-UAV coordination and cooperation, or has studied the human factors issues involved in single-operator supervision of multiple semi-autonomous UAVs. This chapter begins by giving a brief overview of fully autonomous multi-UAV research and then focuses on research involving single-operator human supervisory control of multiple UAVs. This summary highlights the need for development of decision support tools for multi-UAV schedule management. The chapter also introduces configural displays and the advantages of using them for decision support.

2.1. Overview of Fully Autonomous Multi-UAV Research

The majority of multi-UAV research has examined the coordination and cooperation of fully autonomous vehicles with little to no human interaction. Work on completely autonomous multi-UAV control has included path planning [8], cooperative dynamic target tracking [9], task assignment, and mission planning. Task assignment is a particularly challenging area of autonomous multi-UAV control, especially for cooperative UAVs operating in dynamic environments. Proposed methodologies to solve task assignment problems include the use of quasi-decentralized [10] and decentralized [11] task assignment and genetic algorithms [12]. Autonomous UAV research has also examined real-time task allocation with moving targets [13], task assignment under time constraints using Mixed-Integer Linear Programming (MILP) [14] [15], and time-optimal multi-UAV coordination [16]. Multi-UAV mission planning research combines path planning and task assignment with specific mission constraints and objectives, such as the tracking and prosecution of moving ground targets [17].

Some research has considered real-time multi-UAV mission planning in dynamic uncertain environments using a team utility function and system-predictive stochastic model to assign tasks to a networked UAV team [18].

A drawback of autonomous multi-UAV control is that resulting plans and assignments generated from algorithms are often sub-optimal [19]. Additionally, little work has been conducted in utilizing algorithms for real-time mission re-planning due to emergent or unexpected events. Much of fully autonomous multi-UAV research has only examined satisfying time constraints for task assignment and few attempts have delved into real-time mission schedule management. Although fully autonomous multi-UAV research is important, it is likely that one or more human operators will supervise and interact with the UAVs. While these UAVs may utilize some of the capabilities originating from fully autonomous UAV research, the inclusion of an operator in the multi-UAV system will provide dynamic, real-time mission re-planning capability under uncertain and/or emergent situations. An operator may especially be useful for time-critical schedule management, when processing times for algorithms solving complex scheduling problems may take excessive amounts of time. Humans may be better suited to make schedule management decisions, particularly in highly uncertain environments, but they may need assistance from DSTs because of the high cognitive workload involved in multi-UAV supervision.

2.2. Human Factors Research in Multiple UAV Control

In studying single operator supervisory control of multiple UAVs, operator performance, mental workload, and situation awareness are principal concerns. Operator performance is usually characterized by how well the operator achieves mission or task objectives. Mental workload is described as the amount of cognitive capacity applied to one or more tasks. Situation awareness (SA) is generally defined as the perception of elements in an environment within time and space (called level 1 SA), the comprehension of their meaning (level 2 SA), and the projection of their future states (level 3 SA) [20].

Research on single-operator supervision of multiple UAVs has primarily focused on how the above operator characteristics are affected by the number of supervised UAVs [21] [22] [23], by different automation levels [22] [23] [24] [25], and by supervision from manned aircraft [26] [27]. Very few of these studies have examined the temporal component of multi-UAV supervision, and almost none have considered schedule management in time-critical missions. Multi-UAV supervision simulation test beds such as the UAV Modeling and Analysis Simulator Testbed (UMAST) [22], Adaptive Levels of Automation (ALOA) test bed [23], the Multi-Modal Immersive Intelligent Interface for Remote Operation (MIIIRO) [25], and Operator Vehicle Interface (OVI) [28] do not use any kind of timeline representation of the temporal aspects of their missions. Some Multi-UAV simulation interfaces have used a countdown method to represent time pressure in missions [29], but have not provided any way to alleviate high workload situations in advance. Other single-operator, multi-UAV studies have employed tabular-like listings of mission tasks in task schedule windows [30] or event horizons [31], providing some temporal information but little to no capability to manage the mission schedule.

Hanson, et al [32] used a temporal monitor display for time monitoring of individual and team tasks. Part of this display consisted of a timeline showing team tasks for different mission schedule times. This task-based timeline allowed the user to construct, analyze, and monitor mission plans along with the assistance of a "mixed-initiative interaction window" which provided the operator with multiple courses of action. However, the temporal monitor display did not offer the operator any predictive capability about what would happen to the mission schedule if he or she accepted one of the suggested courses of action. Operators therefore had no assistance to help them consider the consequences of their actions.

Cummings & Guerlain [33] implemented timelines in a multiple unmanned Tactical Tomahawk missile supervision simulation. These timelines allowed controllers to "perceive important temporal relationships such as missile launch time, time of impact, and time of fuel remaining, all in comparison to the actual time and to each of the other missiles" [33]. During simulations, operators were tasked with retargeting missiles for specific time on targets (TOTs) and were provided with temporal information on whether the retargeted missile would reach the target on time for its TOT and when it would arrive. Operators could use an interactive decision aid, called the window of opportunity (WOO) to explore future "what-if" possibilities for missile retargeting [34]. The WOO's presentation of "what-if" temporal information corresponding to different decision alternatives is the only example of projective schedule management found in multi-UAV supervision literature, but whether the WOO assisted operators and improved performance was not investigated.

2.2.1. Automated Schedule Management Decision Support Project

The research in this thesis was largely motivated by the work of Cummings & Mitchell [35], who used a graphical timeline with future high workload prediction in a single operator, multi-UAV study. This research examined how varying levels of automation, as represented by different timeline designs, affected operator and system performance in supervising four UAVs in a time-critical targeting mission. Timelines were integrated into a simulation called the Multi-Aerial Unmanned Vehicle Experiment (MAUVE) test bed. Each experimental subject was presented with the MAUVE map display and one of four different timeline display designs.

One of the operator's critical tasks in MAUVE was to arm and fire UAV payloads at targets during their respective TOT window. Because of the time-critical, multiple task nature of the mission, it was possible that several UAVs could require concurrent arming and firing, creating potential high workload periods for the operator. In order to mitigate this overlap of simultaneous targeting tasks, called TOT conflicts, an operator could request a schedule change for a target, called a TOT delay, in order to push the target's TOT into the future [35]. However, these requests were not always granted, as a simulated mission commander approved or denied requests based upon how far in advance they were made. Thus, the earlier operators requested a TOT delay, the more likely it was granted, with near-term requests rarely being approved [35]. No "what if" capability was provided to help operators to understand beforehand the effects of delaying a TOT on the mission schedule. As a result, granted TOT delays could create other future TOT conflicts, or could cause UAVs to arrive late to their assigned targets.

In the Cummings & Mitchell study [35], operator involvement in arming and firing UAV payloads, as well as the timeline decision support, depended upon the assigned automation level. Experimental results unexpectedly showed that human subjects under a management-by-consent automation level had the worst performance in supervising multiple UAVs, regardless of operational tempo [35]. Under this level, operators manually armed and fired UAV payloads, and the timeline graphically showed scheduled UAV actions, used reverse-shading to represent TOT conflicts, and provided recommendations to request target-specific TOT delays to eliminate conflicts. The poor performance caused by this timeline design was traced to its operators misusing TOT delay requests, which should have been used sparingly to manage the mission schedule. Management-by-consent operators were unable to implement effective stopping rules when trying to achieve schedule changes. Instead, they focused more on globally optimizing their schedule and less on performing present mission tasks, which negatively affected their performance and situation awareness [35].

TOT delay request overuse by management-by-consent operators may have been prompted by overly salient representations of advance high workload prediction and poor workload mitigation recommendations, neither of which included any uncertainty information. The reverse shading technique used to notify operators of possible TOT conflicts may have biased them to primarily focus on fixing their schedule, provoking excessive TOT delay requests [35]. Automated recommendations prompting TOT delay requests may have encouraged operators to make excessive requests in order to achieve specific schedule changes. These recommendations did not provide any information about the likelihood of an approved request, or about how TOT delays would affect the overall mission schedule.

It is hypothesized that management-by-consent operators did not fully understand the impact of delayed TOTs on their future mission schedule because they were not explicitly provided with "what if" information to inform their decisions and actions. Schedule changes could create other TOT conflicts later in the mission schedule, or even late arrivals of UAVs to targets. Because no information was provided about the uncertainty involved in requesting TOT delays, operators could not understand the effects of their schedule management decisions. In summary, the management-by-consent decision support timeline was over-salient and did not convey the uncertain effects of schedule changes on the future mission schedule, which contributed to degraded operator and system performance.

Multi-UAV operators need to better understand the potential consequences of schedule management decisions on both current and future schedules. By understanding these effects, operators may generate better stopping rules for schedule optimization, prompting them to only request schedule changes that contribute toward achieving mission objectives. A "what if" capability presenting the consequences of schedule management decisions could inform operator decisions, prompt fewer schedule change requests, and increase operator performance and situation awareness.

Other researchers have also commented upon the need for "what if" predictive capability in single operator control of multiple UAVs [36] [37].

2.3. Configural Displays

Information visualization can be defined as the representation of abstract data by visual elements in order to amplify cognition [38]. Information visualizations present compact graphical representations that can often be perceived and analyzed faster than text-based displays. Fast and efficient perceptual reasoning lays a strong foundation for analysis and decision-making tasks, as information needs to first be perceived before it can be used. Visualizations that do not support good perceptual reasoning can make it difficult for users to gather information needed for tasks, prompting an increase in task performance times, inefficiency, and errors. One type of visualization design, a configural display, especially supports perceptual processes.

A configural display maps several individual variables into a single geometrical form to provide integrated information about the variables [39]. Changes in the individual variables cause the configural display's shape to vary [40], graphically providing dynamic information about changing system properties. In addition, configural displays support the proximity compatibility principle [41] by integrating together the variables needed for comparison or computation.

The goal of configural displays is to support direct perception-action, allowing operators to directly perceive a system state and immediately act upon the gathered information with little contemplation. Support of direct perception-action permits operators to utilize efficient perceptual processes rather than cognitively demanding processes that rely on memory, integration, and inference [42]. Direct perception-action is facilitated by a configural display's emergent features, which are produced by interactions between display elements which represent variables, providing a higher-level aggregate view of a system's state [40]. Use of direct perception-action in user display design has shown improved performance in complex tasks [43] [44].

The benefits of configural displays make them valuable for use in multi-UAV schedule management, an area involving many dynamic variables. In order to address the multi-UAV schedule management issues previously discussed, a configural display was developed to provide operators with current schedule information, as well as a "what if" predictive capability for potential operator-induced schedule changes. This configural display was used as a decision support tool embedded into a slightly redesigned MAUVE simulation to study its effectiveness in improving operator performance, workload, and situation awareness in multi-UAV supervision.

3. Star Visualization (StarVis) Configural Decision Support Tool

This chapter presents the design and implementation of a configural decision support tool called Star Visualization (StarVis), developed for multi-UAV schedule management in time-critical missions. Influence diagram analysis was performed to determine the variables relevant to schedule management, motivating the design of StarVis. StarVis was implemented into the MAUVE multi-UAV real-time mission simulation in two different DST designs, Local StarVis and Quasi-Global (Q-Global) StarVis.

3.1. Influence Diagram Analysis

Initial analysis for the design of a schedule management DST began with the creation of an influence diagram. An influence diagram maps how variables in a system influence one another. The purpose behind creating an influence diagram was to organize and understand (1) what mission variables could influence an operator's attempts to manage a mission schedule and (2) how decision support available to the operator would influence the mission and its schedule. The goal of this effort was to identify those variables that were critical in supporting a human decision maker through the schedule management DST. The influence diagram included the different decision support components, variables related to the multi-UAV mission, and variables not considered in previous schedule management decision support designs [35]. Figure 3-1 shows the influence diagram and the associated legend.

From the influence diagram, it was determined that schedule management decision support should help multi-UAV operators address two different types of schedule problems. First, the decision support should assist operators with mitigating TOT conflicts, which represent a potential high workload period when an operator may need to perform multiple targeting tasks simultaneously. Potentially delaying one of the TOTs in conflict could mitigate the operator's high workload in advance. The second schedule problem uncovered in the influence diagram analysis is called a late arrival, which occurs when a UAV arrives to a target after its scheduled TOT, or if there is not enough time left in the TOT window to execute the targeting sequence. The influence diagram also determined what information about targets should be provided to operators making schedule management decisions. Specifically, operators should know how many targets are involved in the described schedule problems and their relative priorities, so as to understand the severity of specific schedule issues.

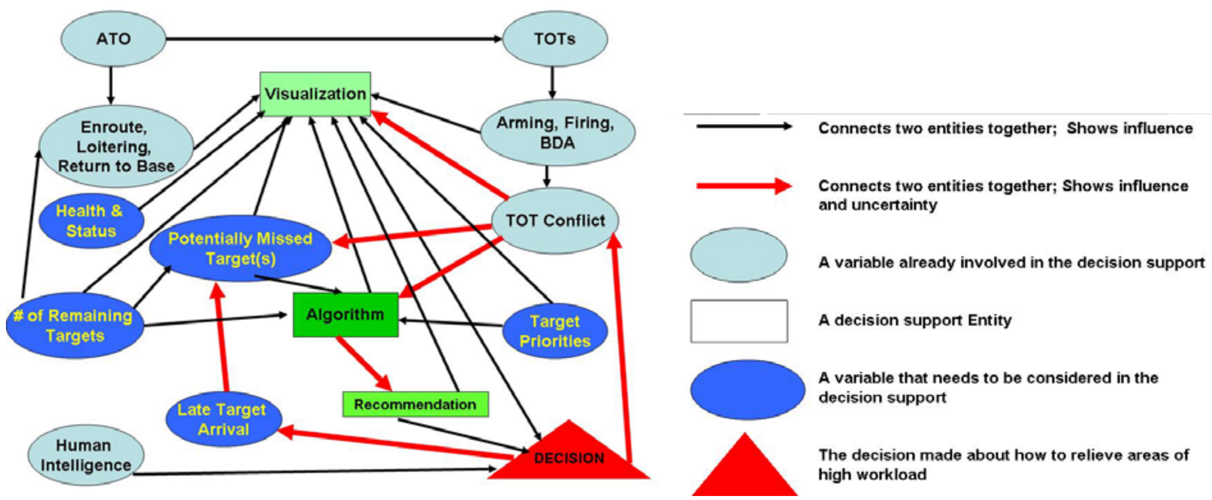


Figure 3-1: Influence diagram for a schedule management decision support tool design

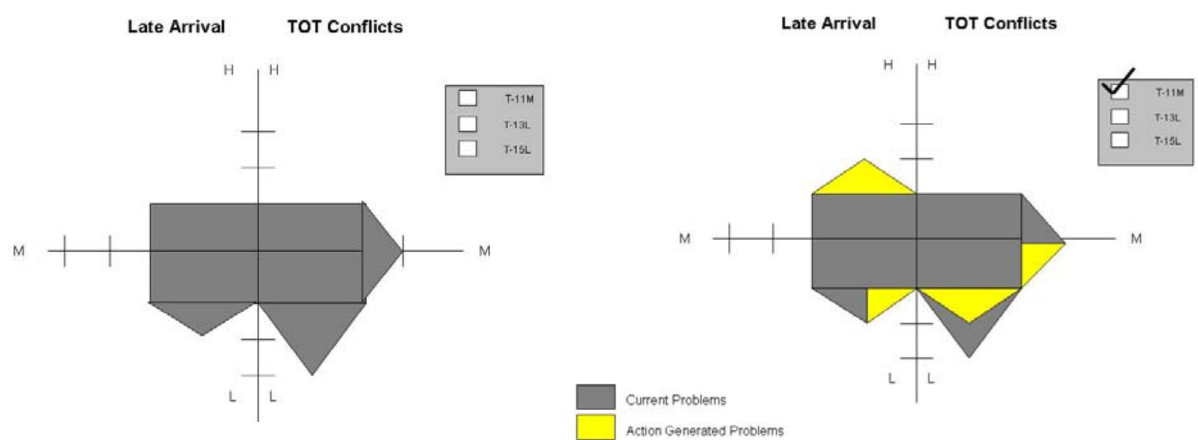
The influence diagram structured the information needed in for the multi-UAV schedule management DST. From this structure, the graphical StarVis configural display was developed.

3.2. StarVis Configural Display

The advantages offered by visualizations make them highly capable and desirable for use as DSTs. Visualizations can comprehensively represent large amounts of information, highlight emergent properties, and facilitate fast and efficient human perception. These capabilities suggest that a graphical DST may be more effective than a text-based DST in helping an operator manage a multi-UAV mission schedule, particularly under time pressure. This was the case for timeline representation in the Cummings & Mitchell study [35], where graphical timeline representations promoted better performance than a text-based timeline. Thus, it was decided to design a graphical schedule management DST in the form of a configural display so as to exploit the benefits of visualizations.

The StarVis configural display, named after its star-like shape, was designed to leverage the benefits of configural displays previously discussed (Section 2.3) to support multiple UAV schedule management. For a single UAV, StarVis represents the number of targets involved in the two types of schedule problems (late arrivals and TOT conflicts), while noting the problem targets' priorities (low, medium, or high). In addition to providing information about targets with current schedule problems, StarVis is a projective "what if" tool, allowing operators to see the potential effects of schedule management decisions projected across the future mission timeline. This predictive "what if" tool was not present in previous decision support designs [35]. Thus, StarVis can display both current and projected future problems for each UAV.

Figure 3-2 shows the StarVis configural display for multi-UAV schedule management. StarVis operates in two modes: current and projected (also called "what if") schedule problems mode. Figure 3-2a shows the default, current problems mode, in which the StarVis indicates schedule problems that currently exist on a single UAV's timeline for the next fifteen minutes. The left side of the StarVis represents targets with late arrivals, while the right side represents targets involved in TOT conflicts. If no problems exist in the next fifteen minutes of a UAV's schedule, StarVis simply displays a gray rectangle. When a UAV's current timeline experiences schedule problems, gray triangles representing those problems grow off its StarVis. High priority targets with a schedule problem (late arrival and/or TOT conflict) are represented by triangles emerging from the top of the rectangle. Targets of medium and low priority are represented by triangles on the sides and bottom of the rectangle, respectively. A triangle's height gives the number of targets of a specific priority involved in a particular schedule problem. In Figure 3-2a, the StarVis shows that for its UAV's schedule, there is one low priority target with an expected late arrival, and one medium and two low priority targets involved in separate TOT conflicts with targets from other UAVs.



(a) (b)

Figure 3-2: The StarVis configural decision support display for multi-UAV schedule management. (a) Current problems mode (b) Projected "what if" mode.

Next to each UAV's StarVis configural display is a list of the UAV's targets that have schedule problems along with selectable checkboxes. When an operator selects a checkbox, the StarVis projective "what if" mode is engaged, as shown in Figure 3-2b. By selecting a checkbox, the operator virtually queries "If I request a TOT delay for this target and it is granted, what could potentially happen to this UAV's schedule?" While the StarVis "what if" mode predicts the potential effects of delaying target TOT(s) on the mission schedule, it does not necessarily indicate exactly what will happen if the selected target TOTs are actually delayed. When in "what if" mode, yellow triangles may possibly, but not necessarily, appear on StarVis, representing new problems that could arise if the selected target is delayed. Split gray and yellow triangles indicate that the current schedule problem could persist if the selected target is delayed. Gray triangles continue to indicate current schedule problems. Therefore, when in "what if" mode, a gray triangle signifies that the schedule problem it represents will potentially be eliminated if the selected target(s) is delayed. For the example shown in Figure 3-2b, if a TOT delay request for the selected target is granted, the UAV could continue to have a late arrival at a low priority target and have a medium priority target involved in a TOT conflict, problems that exist for the current schedule. Additionally, delaying the selected target could decrease the number of low priority targets involved in a TOT conflicts from two to one, and possibly create a new late arrival on a high priority target.

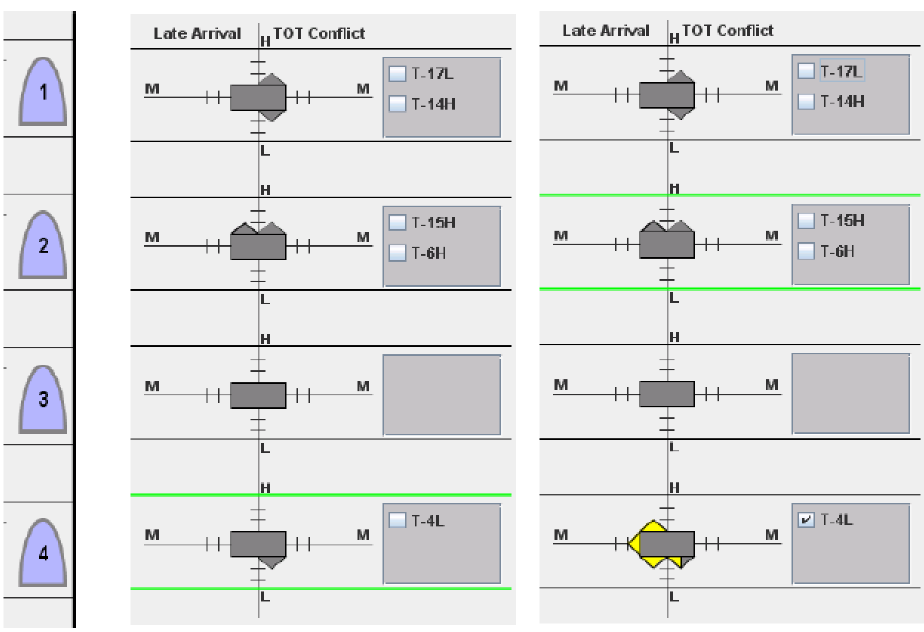
The StarVis configural display supports direct perception-interaction, the use of emergent features, and the proximity compatibility principle. Its graphical form facilitates direct perception of the state of a UAV's schedule, allowing the operator to quickly and easily tell if a UAV has any schedule problems (as indicated by the number of triangles on the rectangle and their heights) or not (as indicated by a rectangle with no triangles). As a mission schedule begins to experience problems, visual representations of these problems "emerge" on the StarVis as triangles grow from the rectangle. By comparing the surface areas of each UAV's StarVis, an operator can quickly discern which UAV is experiencing the most problems and specifically what kind. Thus for schedule management, the StarVis configural display provides a high-level comprehensive overview of problems through emergent features, but also offers low level details on particular variables of interest.

3.3. Implementations

The StarVis configural display was implemented into MAUVE in two different DST designs: A Local StarVis DST and a Quasi-Global (Q-Global) StarVis DST. The two DSTs identically indicate current schedule problems through gray triangles; however, the designs differ in the operation of the projective "what if" mode. For both StarVis DST designs, each UAV has its own StarVis configural display.

3.3.1. Local Implementation

The Local StarVis DST is shown in Figure 3-3. Next to each UAV's StarVis display is a list of targets with schedule problems for that UAV's current schedule, as represented by gray triangles on the StarVis (see Figure 3-3a).



(a) **(b)**

Figure 3-3: Local StarVis DST implementation. (a) Current schedule problems mode (b) Projected "what if" schedule problems mode

In the Local DST design, an operator can only select one target checkbox for each UAV's StarVis in order to activate the projective "what if" mode. When a checkbox is selected, the UAV's StarVis configural display shows the effects of delaying the selected target only on that particular UAV's schedule. Thus, in the Local StarVis DST, yellow "what if" triangles can only appear on an individual UAV's StarVis if a target checkbox belonging to it is selected. Notice that in Figure 3-3b, UAV 4 is the only StarVis with a selected checkbox and thus only yellow triangles appear on its StarVis. Although each StarVis may have only one target checkbox selected at a time, more than one StarVis may have a checkbox selected. For example, the StarVis displays belonging to UAVs 1 and 3 may each have yellow triangles if each UAV has one checkbox selected. However, when in "what if" mode, a UAV's StarVis only uses the other UAVs' current schedule information in its projection, and does not consider any of their StarVis' "what if" information. When multiple StarVis's are in "what if" mode, operators may compare decision alternatives to resolve schedule problems, particularly for TOT conflicts.

3.3.2. Quasi-Global (Q-Global) Implementation

The Q-Global StarVis DST, shown in Figure 3-4, is termed "quasi-global" because it's "what if" mode allows operators to explore solutions to multiple schedule problems, showing the effects of potential schedule management decisions across all UAV timelines. Q-Global StarVis was designed in response to the Local StarVis's inability to show schedule management decision consequences across all UAVs. The Q-Global design differs from the Local StarVis in that all targets with current schedule problems are listed together to the right of the StarVis displays (as seen in Figure 3-4a), instead of separated by their respective assigned UAVs. If no target checkboxes are selected, the Q-Global StarVis displays exactly the same as the information as the Local StarVis with unselected checkboxes. The two DST designs only differ when checkboxes are selected, engaging the "what if" mode.

Multiple checkboxes may be selected in the Q-Global StarVis DST to show what could happen to ALL UAV schedules if the selected target TOTs are delayed. When one or more checkboxes are selected, the projected problems (if they exist) appear across all the StarVis configural displays. In Figure 3-4b, one target checkbox is selected and "what if" information in the form of yellow and split appears on multiple StarVis displays. In contrast to the Local design, Q-Global StarVis shows the effects of TOT delays on the entire mission schedule, instead of only on the UAV schedule the target is assigned to. Additionally, because Q-Global StarVis allows for selection of multiple target checkboxes, operators may layer solutions to multiple schedule management decisions.

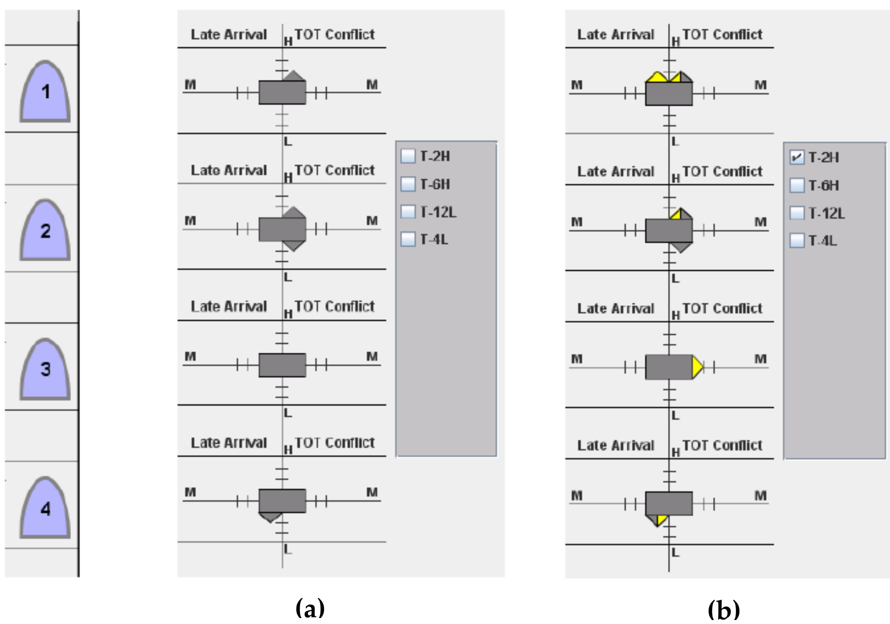


Figure 3-4: Q-Global StarVis DST implementation. (a) Current schedule problems mode (b) Projected "what if" schedule problems mode

3.4. Summary

The purpose of the StarVis configural DSTs is to help operators manage a multi-UAV mission schedule by providing (1) comprehensive information about current schedule problems, and (2) "what if" information about how TOT delays could either help or hurt future mission schedules. StarVis' configural form offers the advantage of direct perception-action through its emergent features and supports the proximity-compatibility principle, making it easy for operators to quickly assess the schedule's condition and make management decisions.

The Local StarVis DST shows operators the effects of one schedule change on one UAV's timeline, allowing operators to directly compare the consequences of different decision alternatives for a single schedule problem. This is particularly useful in fixing TOT conflicts; operators can select the targets involved in the conflict and directly assess how delaying one of the targets could affect each UAV's schedule. In contrast, the Q-Global StarVis DST shows operators the aggregated effects of multiple schedule changes on all UAVs, providing greater system-wide understanding on how schedule alterations affect the whole multi-UAV mission, as opposed to just one UAV. The Q-Global StarVis also allows for decision-layering, permitting operators to see how more than one TOT delay could affect the overall mission schedule. The advantages and drawbacks of both StarVis DST designs motivated the need for human subject testing, which is described in the next chapter.

4. Methods

A human subject experiment was conducted using the MAUVE simulation and the two implementations of the StarVis configural DST. The experiment measured performance, workload, and situation awareness of operators using Local StarVis, Q-Global StarVis, or no schedule management decision support. This chapter describes the experimental question, subjects, apparatus, tasks, and design used in this study.

4.1. Experimental Question

Due to the difficulty humans have in accurately predicting the effects of schedule management decisions in the face of uncertainty [35], it was hypothesized that operators would better manage a multi-UAV mission with either one of the StarVis DSTs than with only a timeline. This was hypothesized because of the StarVis configural display's perceptual-based overview of all UAV schedule problems, as well as its interactive "what if" mode allowing visualization of potential consequences to schedule management decisions.

Because of the benefits and drawbacks of both StarVis implementations, it was not hypothesized which StarVis DST would help operators better supervise the multi-UAV mission. Because operators using the Local StarVis could directly compare multiple alternatives of one schedule management decision, it was predicted that they would be able to mitigate schedule problems for individual UAVs. However, because Local StarVis only displayed decision consequences for one UAV, it was speculated that operators might not consider the consequences of decision solutions across the overall mission schedule. This focus on individual UAV optimization could lead to increases in schedule problems for the overall mission, which could increase operator workload. Because the Local StarVis did not provide operators with "what if" information about how schedule changes affected all UAVs, they might not understand that solutions improving one UAV's schedule could degrade the overall mission schedule.

Because Q-Global StarVis visualized the effects of potential schedule changes across all UAVs in the mission, it was hypothesized that operators could understand how schedule changes affected the overall mission, establishing a system-wide mission perspective. Additionally, Q-Global StarVis could represent a combination of solutions to multiple schedule problems, which could increase operator awareness about how multiple schedule changes affected the overall mission schedule. However, because the Q-Global StarVis did not directly compare the alternatives of one decision, it was theorized that Q-Global operators could spend excessive time optimizing one decision, leading to performance decreases from problem-solving fixation.

4.2. Subjects

A total of 15 subjects, 11 males and 4 females, took part in the experiment. The subject population consisted of students, both undergraduates and graduates, and young professionals in technical fields. All subjects were paid $10 per hour for their participation, and a $50 gift certificate was offered as an incentive prize to the best performer in the experiment. Subject age ranged from 20 to 31 years, with a mean of 24 years. No subjects had any military experience. Three subjects had experience with remotely piloted vehicles (RPVs), with an average of 53 hours between them. Six subjects, including the three subjects with RPV experience, had aircraft piloting experience, either in powered aircraft or gliders. The number of flight hours among this group ranged from 1 to 200 with an average of 93.5. A more detailed summary of subject demographics is provided in Appendix A.

4.3. Test Bed

4.3.1. Experimental Apparatus

The experiment was performed on a four screen system called the multi-modal workstation (MMWS), shown in Figure 4-1.

Figure 4-1: The Multi-Modal Workstation (MMWS)

The top three 21 inch screens were run at 1280 x 1024 pixels, 16-bit color resolution, while the 15 inch bottom screen was run at 1024 x 768 pixels, 32-bit color resolution. The workstation computer was a Dell Optiplex GX280 with a Pentium 4 processor and an Appian Jeronimo Pro 4-Port graphics card. Experimental subjects interacted with the MAUVE simulation through a generic corded mouse and cordless keyboard. The top leftmost screen contained a listing of the mission objectives in priority order for the scenarios and was static throughout the entire experiment. The top middle screen contained the MAUVE map display and the top rightmost screen contained one of three different MAUVE decision support timeline displays, which are described in the next sub-section. During training, the bottom middle screen contained a color coded reference table for UAV actions in the simulation. Appendix B shows the mission objectives and color coded reference screens.

4.3.2. Experimental Task: The Multi-Aerial Unmanned Vehicle Experiment (MAUVE)

The previously described MAUVE simulation [35] was used in the experiment with some modifications. MAUVE presented pre-planned missions to the operator for real-time execution. During the simulation, operators were tasked with arming and firing UAV payloads at scheduled times, re-planning UAV paths in response to emergent threats, assigning emergent targets to the most appropriate UAV, and answering questions about the mission from an automated "supervisor" through the instant messaging window.

Figure 4-2 shows the map display that all experimental subjects used. This display provided mission time information, a geo-spatial representation of the UAV flight paths (black or green lines), waypoints (black triangles), targets (red diamonds), threat areas (yellow circles), and a mission planning and execution toolbar for operator interaction with the UAVs. Each UAV had its own mission planning and execution toolbar which was active only when a UAV was selected, highlighting its path in green. Operators used the toolbar to arm and fire UAV payloads at targets of different priority (low, medium, and high), command the UAV to move to the next assigned target or return to base, assign or un-assign targets to a UAV, rearrange the order in which a UAV visited targets, request TOT delays for targets, schedule or unscheduled battle damage assessment (BDA), and add and remove waypoints.

The only alteration of the map display between the MAUVE used in previous experiments [35] and this experiment was the addition of a qualitative TOT delay request probability bar placed above the "Req[uest] TOT Delay" button. Depending on how far into the future a selected target was in a UAV's schedule, the probability bar displayed the likelihood of a TOT delay request approval for that target. Target TOTs scheduled in the next five minutes were given a low probability of being delayed, those between five and ten minutes into the future had a medium probability, and TOTs ten to fifteen minutes away had a high probability of an approved request. The probability bar was positioned above the TOT Delay Request button so as to inform subjects of the likelihood of request approval before they requested a delay for a particular target. Subjects could therefore directly consider the feasibility of actually achieving their decision solutions, and thus potentially generate better stopping rules when trying to achieve particular delays.

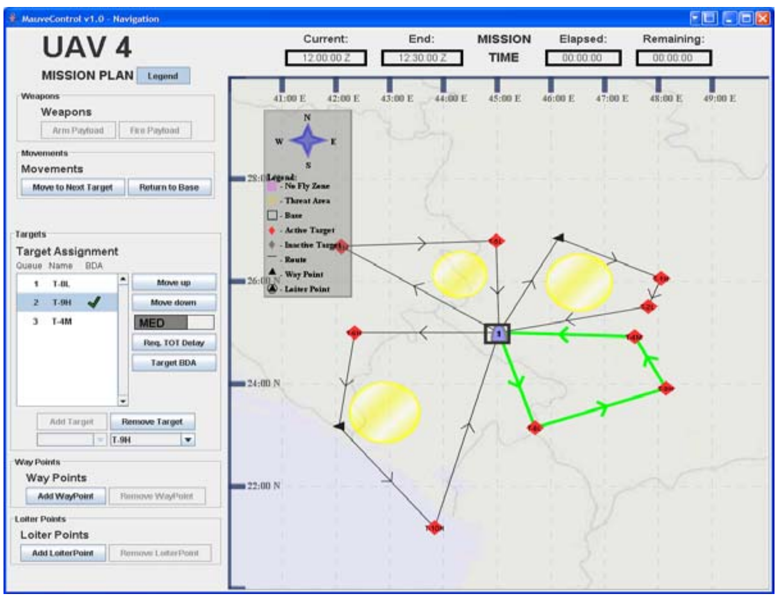


Figure 4-2: Map display of the MAUVE interface

In addition to the map display, subjects used one of three different decision support timeline designs, all of which were identical except for the schedule management DST they included. Decision support timeline displays contained UAV status information, a color-coded graphical timeline, either one of the StarVis DSTs or no DST, an instant messaging window for human to human communications, and a

UAV data link window for UAV to human communications. The graphical timeline represented all four UAV schedules for the next fifteen minutes with colored bars representing different UAV flight phases, as shown in Table 4-1. A TOT was defined as the arming (yellow), firing (red), and BDA (brown) windows on the timeline. In order to arm and fire, a UAV had to be flying over a target during its TOT window. UAV payloads could be armed during the arming and firing windows, but could only be fired during the firing window. Figure 4-3 shows the Local StarVis decision support timeline design as an example.

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

Table 4-1: UAV color-coded flight phases.

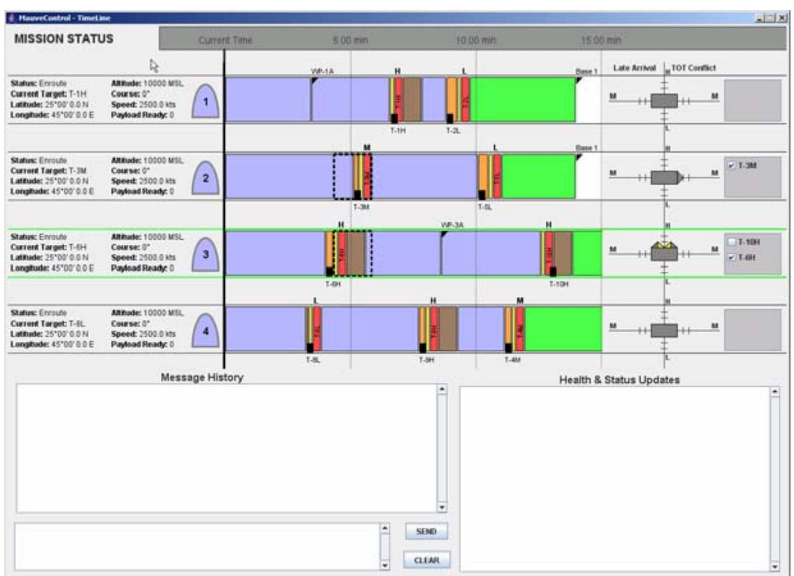
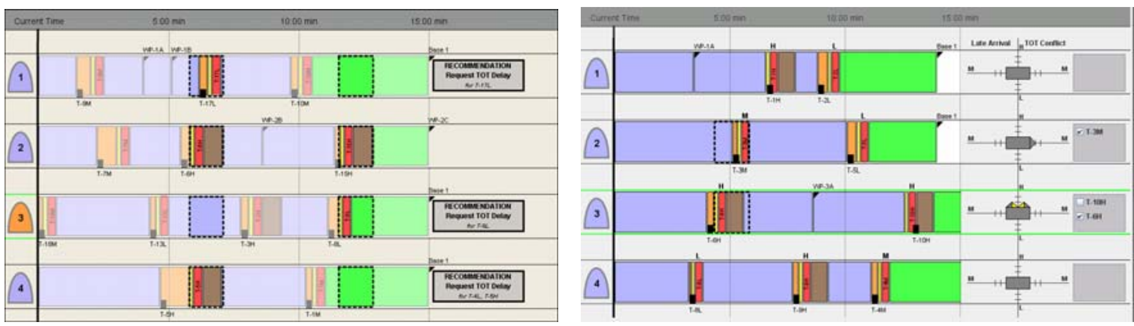


Figure 4-3: Example MAUVE decision support timeline design – Local StarVis design

The experiment used the management-by-consent automation level from previous work [35], which was altered by replacing the automated recommendation with either the Local StarVis, Q-Global StarVis, or no DST. The reverse shading used to indicate TOT conflicts was eliminated. Instead, potential high workload periods were denoted by dashed boxes surrounding those targets with conflicting TOTs. Figure 4-4 shows the original management-by-consent [35] and redesigned color-coded timelines.



(a) **(b)**

Figure 4-4: Change in MAUVE management-by-consent timeline design. (a) Original design [35] (b) Redesigned timeline.

UAV arrival to a target was represented on the timeline by a small black box labeled with the target's designation. If the UAV arrived to the target early (before its TOT), this box was located to the left of the target's respective TOT window, or within the window. As described in Chapter 2, UAVs could arrive late to targets, after their scheduled TOT. A late arrival was indicated on the timeline by the UAV target arrival box appearing to the right of the target's respective TOT (after the window). Figure 4-5 shows an example timeline, highlighting a UAV target arrival box, an example late arrival, and a TOT conflict.

In the experiment, a subject's main task was to supervise four UAVs in a time-critical targeting mission as provided by MAUVE. Specifically, the subject's primary objectives were to guide each individual UAV's actions so that all UAVs properly executed required mission commands, and correctly re-plan the mission based upon emergent events. The operator's secondary objective was to answer mission status questions through the instant messaging tool. Supervision of the entire mission was broken down into prioritized sub-tasks, listed from highest priority to lowest:

1. Return to base (RTB) within the mission time limit
2. Obey changing mission requirements as relayed by intelligence messages
3. Destroy all targets before the end of their time on target (TOT) window by manually arming and firing a weapon
4. Perform battle damage assessment (BDA) on specified targets after destroying them
5. Avoid damage from enemy fire by navigating around and out of threat areas
6. Answer communication questions

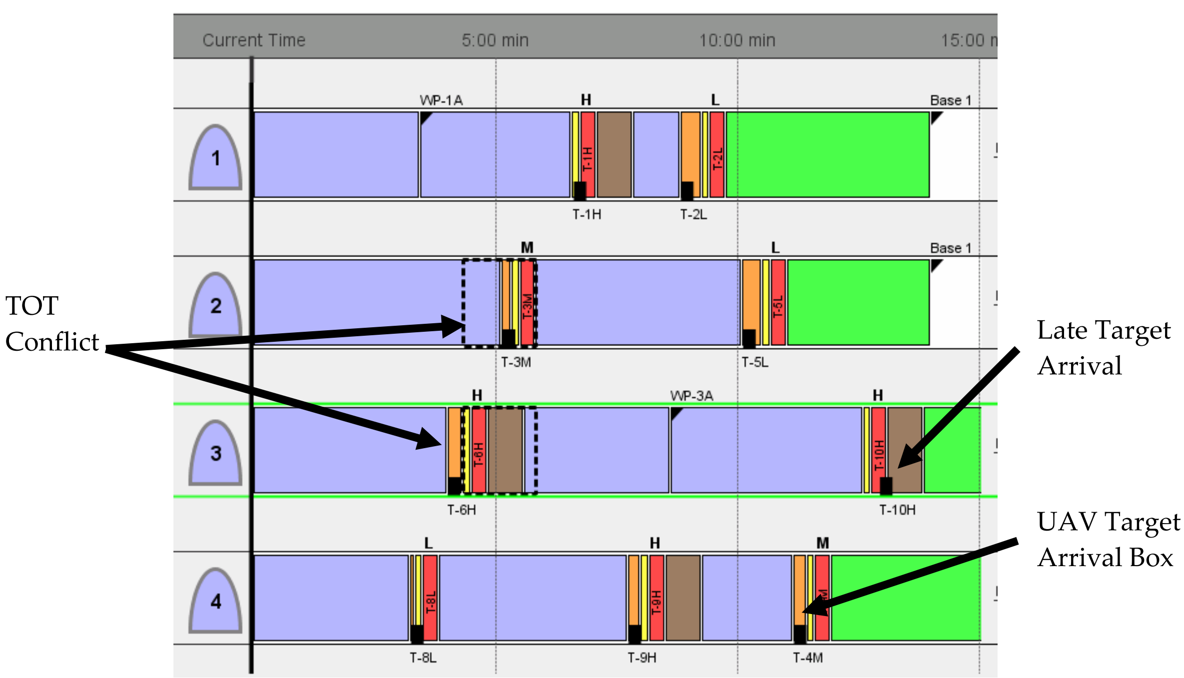


Figure 4-5: Redesigned timeline highlighting TOT conflict and late target arrival schedule problems.

Operators were explicitly trained to follow this priority list, which was displayed throughout training and test sessions. In supervising the mission, the operator needed to manage the mission schedule in order to achieve mission objectives. This involved executing the schedule and addressing predicted late arrivals and TOT conflicts. Late arrivals could be fixed either by path re-planning, if possible, or by requesting a TOT delay. TOT conflicts could only be mitigated by requesting a TOT delay. The operator could used the timeline and StarVis DST (if provided) to explore different decision alternatives for solving specific problems, and to manage the overall mission schedule.

4.4. Experimental Design

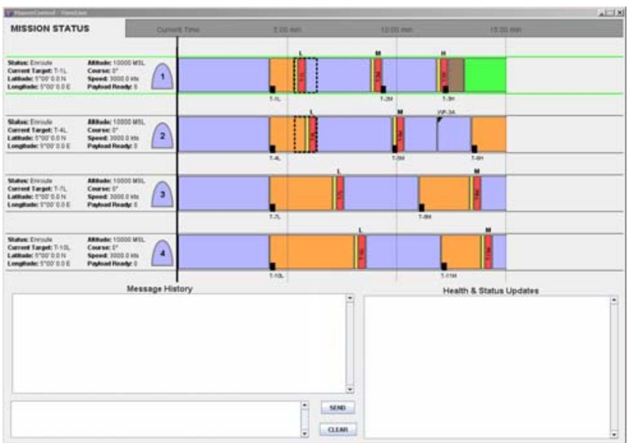
The experiment was a 3 (decision support tool) x 2 (level of re-planning) mixed design study. The level of re-planning factor was within-subjects, and the decision support timeline factor was between-subjects.

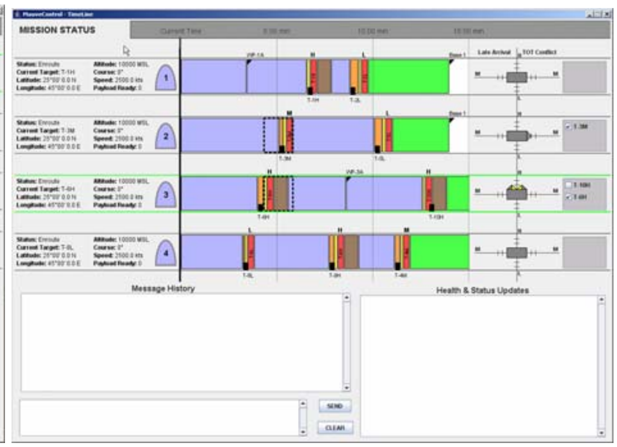
4.4.1. Independent Variables

Two independent variables were of interest in this experiment: schedule management DST and level of re-planning. There were three schedule management DSTs as represented by different MAUVE decision support timelines: no DST, Local StarVis DST, and Q-Global StarVis DST, shown in Figure 4-6. Each experimental subject used one of these three decision support timeline displays.

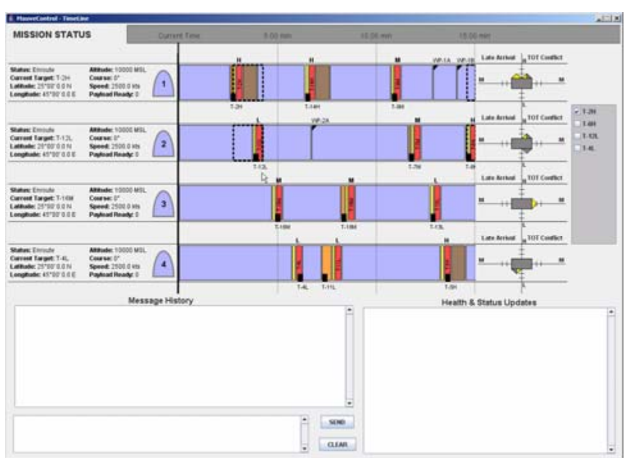
The second independent variable, level of re-planning, represented the operational tempo of re-planning events in a mission scenario, both in the number of events and how they were spaced. This variable was included in the study to examine operator performance and situation awareness under different workload levels. Types of re-planning events included:

- Emergent targets which needed to be added to a UAV's schedule
- Re-assignment of a target to a different UAV strike mission
- Appearance of a new threat area in the map display
- Disappearance of a threat area
- The addition of BDA to an existing target's schedule
- The removal of BDA from an existing target's schedule
- A command for a UAV to return to base during the mission





(a) **(b)**



(c)

Figure 4-6: The three experimental decision support timeline displays. (a) No DST (b) Local StarVis DST (c) Q-Global StarVis DST

Each subject was exposed to a low and high re-planning level in two separate mission scenarios. The low re-planning level contained 7 events spaced by approximately 3 minute intervals, with each interval containing only one event. The high re-planning level contained 13 events spaced at approximately 3 minute intervals, with each interval containing 2-3 events.

4.4.2. Dependent Variables

Several dependent variables were studied in this experiment: a performance score, the number of TOT delay requests, the percentage of approved TOT delay requests, secondary and subjective workload measures, situation awareness, the number of critical firing events, late arrival and TOT conflict mitigation scores, and DST interaction.

Performance Score

In order to measure subject performance for each mission scenario, a performance score was adapted from a previously designed score used in a similar experiment [45]. A subject's performance score was based upon the total number and type of mission objectives completed over an entire testing session, with penalties applied for actions which could result in negative consequences for the mission. Subjects could earn points by correctly destroying targets, which had different values based upon their priority and difficulty level, and by correctly performing BDA for certain targets. Subjects incurred performance score penalties for incorrectly firing at targets specified not to be destroyed, allowing UAVs to traverse threat areas, and permitting UAVs to operate beyond the mission time limit of thirty minutes. The performance score also penalized operators for TOT delay requests, as abuse of this capability could have negative consequences for an individual and organization in actual time-critical military operations. Thus, the performance score was composed of earned (positive) and penalized (negative) points, with higher scores indicating better performance. Appendix C gives more details on the performance score formula.

Number of TOT Delay Requests

The number of TOT delay requests was measured in order to verify if a negative correlation between performance and excessive requests existed. Previous research found that under a management-by-consent level of decision support, the number of TOT delay requests was higher and performance was lower when compared to other decision support designs [35]. The number of TOT delay requests was simply a count of the requests an operator made in each mission scenario.

Percentage of Approved TOT Delay Requests

The percentage of approved TOT delay requests was measured to see if either StarVis DST helped operators manage their schedule farther in advance. As the probability of an approved request for a target decreased the closer its TOT was to the present time, a difference in this metric across DSTs would indicate that subjects using different DSTs managed different sections of the schedule. If statistically significant, this metric could signify that a StarVis DST prompted subjects to focus on certain parts of the mission schedule.

Secondary Workload

Workload measures were relevant to this experiment as the StarVis DSTs were intended to help operators alleviate potential high workload periods (TOT conflicts) in a mission schedule. Secondary workload was measured by the average response time to online instant message questions that appeared at predetermined times in each mission scenario. Previous research showed that the use of instant message questions is an effective technique for measuring spare mental capacity, and thus workload, in command and control settings [46]. The response times to all questions were averaged over the total number of questions asked in a mission scenario. When an operator did not answer a question, the response length was taken as the time between when the question was asked to when the next question appeared. It was assumed that if questions remained unanswered, the subject was experiencing high workload, since answering communications was the lowest priority mission objective (see Appendix B).

Subjective Workload

In addition to reducing actual workload, addition of a workload mitigation DST should not increase perceived operator workload. Operator subjective workload was measured using the NASA Task Load Index (TLX) subjective workload rating survey. A workload score was calculated from operator-weighted ratings on a 1-20 scale along dimensions of mental demand, physical demand, temporal demand, effort, performance, and frustration [47]. Because the mission task involved no physical demand, subjects were instructed to purposefully rank physical demand as a low contributor to workload and ignore survey portions asking about that dimension. Thus the NASA TLX survey was modified to measure the subjective cognitive workload of a subject supervising multiple UAVs.

Situation Awareness

A combined measure of level 2 and 3 situation awareness (SA) was adapted for this experiment based on previous measures [45]. Four indicators of situation awareness were determined from mission scenarios:

1. The number of significant entries into threat areas, defined as when a UAV received more than 3 hits and the operator took no action to minimize further hits.
2. The amount of time UAVs loitered at missed targets due to loss of SA, as observed by the investigator.
3. The number of targets missed due to lack of SA, as observed by the investigator.
4. The percentage of re-planning events successfully and correctly completed.

These four indicators combined represented level 2 (comprehension) and level 3 (future projection) situation awareness [48]. Different ranges of possible values for each of the SA indicators were grouped and then ranked on a 1-5 scale. Table 4-2 shows the 1-5 scale and the relative range of values for each indicator [45]. For each mission scenario, subjects were ranked according to the scale for each of the SA indicators, and a total SA score was determined by averaging the SA indicator scores.

Situation Awareness Score	Number of significant entries into threat areas, no operator intervention	Amount of time UAVs spent loitering at missed or removed targets (seconds)	Number of targets missed due to lack of SA	Percentage of re-plans successfully completed
5	0	0 - 30	0 -1	90 or more
4	-	30-90	2 - 3	80 - 90
3	1	90-120	4	70 -80
2	-	120-200	5	60 -70
1	2 or more	200 or more	6 or more	60 or less

Table 4-2: Situation awareness indicators and relative scales [45].

Number of Critical Firing Events

The number of critical firing events measured how many times operators incorrectly fired on targets. The command to not destroy specific targets was a message provided to subjects within the mission scenario via the instant messaging tool, and represented a re-planning event. Because incorrect target destruction is generally rare among subjects, the number of critical firing events was summed for all operators using the same DST (either no DST, Local StarVis, or Q-Global StarVis), thus reflecting a total across all subjects as opposed to scores for individual subjects. This metric was an overall global situation awareness measure that signified a critical metric in real-life UAV applications.

Late Arrival Mitigation

The late arrival mitigation score documented the number of late target arrivals the operator either eliminated or created within a mission scenario. This metric measured how well the StarVis DSTs helped operators manage their mission schedule by addressing this particular schedule problem. Additionally, as decreasing the number of late arrivals in the schedule would increase the number of targets an operator could potentially destroy, it was possible that this metric could be correlated with performance. Each time a late target arrival was generated, either through the scenario design or a subject's actions, a point was deducted from the score. When a subject eliminated a late target arrival through proactive schedule management, a point was added. Thus, good performance in mitigating late arrivals was indicated by a high score, the highest of which was 0, indicating that all late arrivals were mitigated. The more negative a late arrival mitigation score was, the less the subject addressed late target arrival problems.

TOT Conflict Mitigation

Similar to the late arrival mitigation score, the TOT conflict mitigation score measured the number of TOT conflicts the operator either eliminated or generated within a mission scenario. This metric addressed how the StarVis DSTs helped subjects reduce the number of potential high workload areas, thus decreasing possible future workload and perhaps increasing performance. For this score, each time a TOT conflict was created, either through the scenario's design or subject's actions, a point was deducted. When a subject eliminated a TOT conflict a point was added. Good performance in TOT conflict mitigation was indicated by a high score. As with the late arrival mitigation score, the highest TOT conflict mitigation score a subject could receive was 0, and the more negative the score, the less the subject eliminated TOT conflicts.

DST Interaction

The amount of interaction subjects using either StarVis DST had with its "what if" mode was calculated by counting how many times subjects either selected or deselected StarVis target checkboxes. Although both StarVis DSTs employed the same configural display and identically visualized current schedule problems, their "what if" capability was very different and could prompt different schedule management strategies and actions. DST interaction was measured to see if Local and Q-Global StarVis subjects used the "what if" mode with a similar or different frequency.

4.5. Testing Procedure

Before arriving to the experiment session, subjects were emailed a pre-experiment PowerPoint tutorial to familiarize them with the MAUVE interface. Tutorials were created for each schedule management DST (No DST, Local DST, Q-Global DST) in order to expose subjects only to the particular experimental condition they would encounter. Upon arrival to the experiment session, subjects filled out an informed consent form and a demographic survey, which is provided in Appendix D. Subjects then read a short introduction sheet detailing the experiment format, mission objectives, example communications, rules of engagement, and details on the configural DST if subjects would be provided with one. Similar to the pre-experiment tutorial, each introduction sheet was customized for each schedule management DST design. Appendix E contains an example of the introduction sheet given to subjects using Local StarVis.

The remainder of the experiment consisted of three distinct phases: training scenarios, experiment scenarios, and post-experiment feedback. In the training phase, all subjects received between 90 and 120 minutes of training over three to four practice scenarios until they demonstrated basic competency in using the MAUVE simulation and achieving mission objectives. Practice scenarios were presented to subjects in the same format and order. The first scenario familiarized subjects with the basic displays, mission execution actions, and rules of engagement, while the second scenario introduced all possible mission re-planning events. Subjects were guided through the first two scenarios by the experiment investigator who used a script. The third scenario consisted of a hands-off 15 minute test similar to the experimental mission scenarios, but also included previously unseen instant message questions about the mission. Passing criteria for this scenario included correctly arming and firing upon a certain number of targets, as well as successfully completing a specific number of re-planning tasks. If a subject did not pass the third training scenario, he or she was allowed to test again on a fourth scenario unique from the previous three. If unable to pass the fourth scenario, a subject was excused from the experiment and compensated for his or her time. However, all subjects in the experiment passed either the third or fourth training scenario.

If the subject demonstrated proficiency in training, he or she was then tested on two consecutive 30 minute mission scenarios, one low and one high mission re-planning level. Each of these scenarios represented a pre-planned mission developed by a separate agency, which is typical of military operations. The low re-planning scenario contained 7 re-planning events, while the high re-planning scenario contained 13. The order in which the subject was exposed to the re-planning levels was randomized and counter-balanced. All subjects experienced the same scenarios, but each subject was provided with only one type of schedule management DST. After each mission scenario, subjects completed a NASA TLX survey. After completion of the last mission scenarios, subjects provided feedback on the MAUVE interface and their schedule management DST by completing a post-experiment feedback questionnaire, presented in Appendix F.

4.6. Data Collection

During testing, all UAV actions, subject mouse clicks, and message box histories were recorded by the MAUVE software. In addition, Camtasia® screen capture software video-recorded both the map and timeline decision support displays.

5. Results and Discussion

This chapter presents the statistical results of the experiment described in Chapter 4 and discusses the key findings, relating them to the research questions. Descriptive statistics for all dependent variables are provided in Appendix A.

5.1. Overview

Because the experiment examined two independent factors, 3 schedule management DSTs (between-subjects) and 2 levels of re-planning (within-subjects), the general linear statistical model used for analysis was a 2x3 repeated measures MANOVA. Five subjects were nested within each schedule management DST, and both independent factors were considered fixed while subjects were a random factor. All dependent variables met normality and homogeneity assumptions, and for all reported results $\alpha = 0.05$, unless otherwise noted. Tukey post hoc comparisons were used to examine dependent variables that were statistically significant across schedule management DST. Additionally, Pearson correlations were calculated to find relevant relationships between performance and the other dependent variables.

5.2. Performance Score

As discussed in Section 4.4.2 and Appendix C, subject performance score was a function of the number and priority of targets correctly destroyed, properly performed BDA, incorrect firing at targets, threat area traversal, mission completion within allotted time, and TOT delay requests. Figure 5-1 shows a box plot of the performance score for the three schedule management DST conditions under the two levels of re-planning.

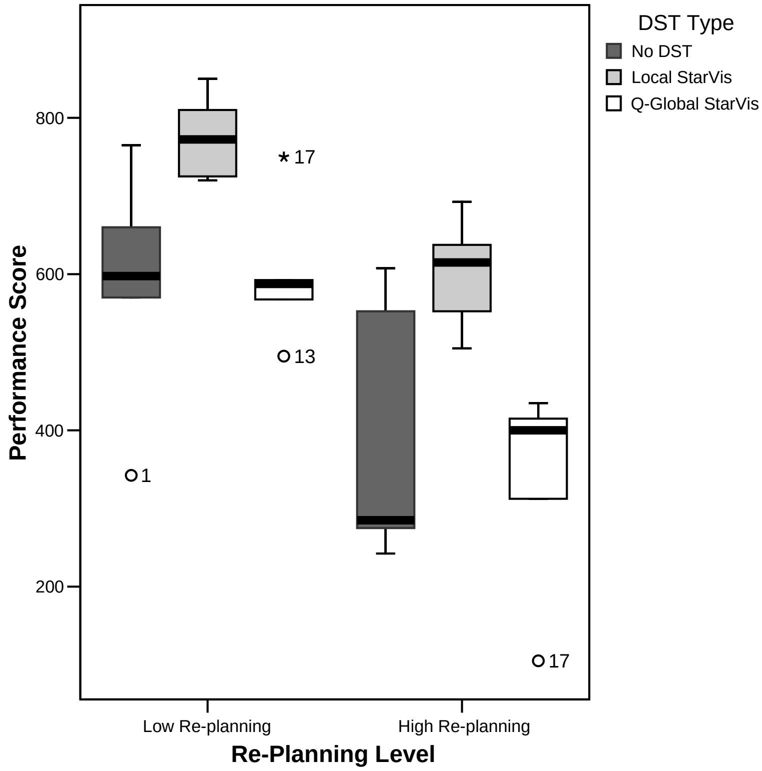


Figure 5-1: Box plot of performance score.

Both the level of re-planning (F(1,12)=22.5, p<0.001) and schedule management DST type (F(2,12)=9.9, p=0.003) were statistically significant. There was no significant interaction between the independent variables. Subjects using the Local StarVis performed better than subjects using no DST or the Q-Global StarVis. Tukey post hoc comparisons determined that Q-Global StarVis subjects performed statistically the same as subjects using no DST (p=0.903), and Local StarVis subjects outperformed subjects using the Q-Global StarVis (p=0.004) and no DST (p=0.009). Furthermore, subjects using either StarVis DST tended to have more consistent performance under high re-planning than subjects using only the timeline. However, under high re-planning, Local StarVis subjects performed almost two times better than Q-Global subjects (Local mean=600.5, Local standard deviation = 73.3; Q-Global mean=333.5, Q-Global standard deviation = 136.0).

It was expected that both StarVis DSTs would facilitate better performance than no DST, but it was surprising that Local StarVis subjects performed the best overall. Subjects who used the Q-Global StarVis performed statistically at the same degraded level as subjects without a DST. When considering average performance across both mission scenarios, some subjects with only a timeline even performed better than Q-Global subjects. Averaged across the entire experiment, the Local StarVis subject with the worst performance still achieved a higher performance score than the Q-Global subject who performed the best. In summary, the Local StarVis improved performance in the multi-UAV supervision task, but the Q-Global StarVis did not, even though both DSTs used the same basic configural display.

5.3. TOT Delay Requests

5.3.1. Number of TOT Delay Requests

The number of TOT delay requests was not significant across schedule management DST type, and was only marginally significant across re-planning level ($F(1,12)=4.1$, $p=0.065$). However, this metric demonstrated a marginally significant interaction between the two independent factors ($F(2,12)=3.8$ $p=0.053$). As shown in Figure 5-2, as operational tempo increased, subjects using either StarVis DST responded by requesting more TOT delays, while subjects without a DST decreased their TOT delay requests, thus causing the interaction.

Subjects using no decision support on average requested the most TOT delays, particularly under low re-planning. Both StarVis DSTs produced the lowest number of requests under low re-planning (a contrast with no DST yielded $t(12)=6.09$, $p<0.001$), demonstrating that under low workload, subjects using either StarVis effectively minimized their TOT delay requests. This minimization signifies that StarVis subjects did not fixate on managing the mission schedule and were able to balance schedule management with their other mission tasks. However, as Figure 5-2 shows, under high re-planning all operators made about the same number of TOT delay requests, regardless of which DST they were assigned.

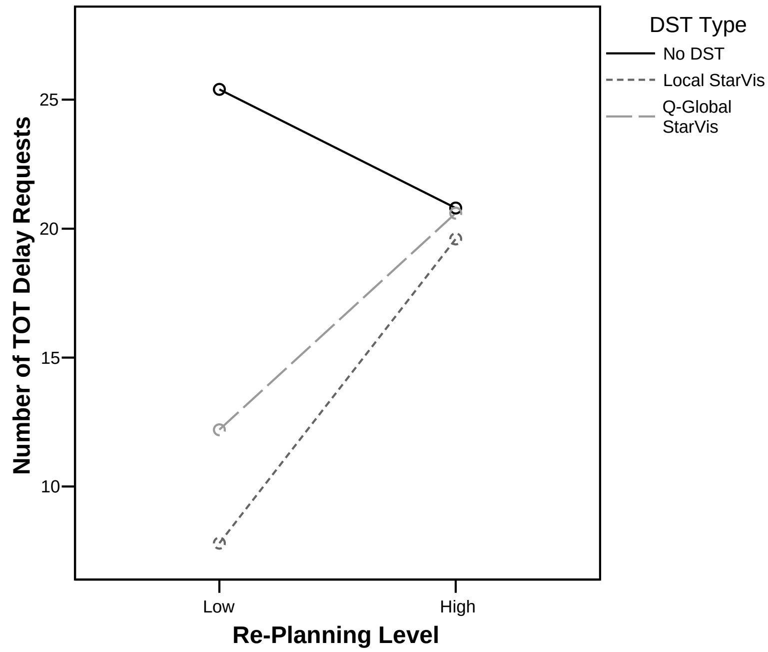


Figure 5-2: Estimated marginal means plot showing interaction for number of TOT delay requests

Under low re-planning, subjects without a DST may have had more time to identify schedule problems from the timeline. However, because they were unable to preview the potential consequences of TOT delay requests on their schedule, they requested more TOT delays than StarVis subjects. Under low re-planning, StarVis subjects perhaps had more time to examine the possible consequences of schedule management decisions and thus requested fewer TOT delays. All this changed under high re-planning, when under time pressure subjects without a DST were unable to quickly identify schedule problems from the timeline, prompting them to request fewer TOT delays. Perhaps under a faster operational tempo, StarVis subjects did not have enough time to examine StarVis "what if" information and thus requested more TOT delays.

Previous research found that when subjects requested many TOT delays, their performance was lower, even when performance scores did not penalize for requests [35]. An overall correlation in this experiment, controlling for the type of schedule management DST, confirmed this earlier finding ($r=-0.581$, $p=0.001$). The more subjects requested TOT delays, the more likely they earned a low performance score. Even though under high workload all subjects virtually requested the same number of TOT delays, operators using the Local StarVis still performed the best overall. When the statistical model for performance included number of TOT delay requests as a covariate, performance score results essentially remained the same, reaffirming that subjects without a DST and Q-Global subjects performed no differently ($p=0.454$), and Local StarVis performance was higher than the no DST and Q-Global conditions ($p=0.039$ and $p=0.001$ respectively).

5.3.2. Percentage of Approved TOT Delay Requests

The percentage of approved TOT delay requests was not significantly different across re-planning level or schedule management DST type, and there was no significant interaction between the independent variables. On average, subjects using either StarVis DST did not have more TOT delay requests approved than subjects using only the timeline. Because approval likelihood was based upon how far in advance a request was made, these results show that statistically, StarVis subjects did not request TOT delays for targets earlier or later than subjects with no DST. Additionally, subjects did not change when they requested TOT delays because of increasing workload.

5.4. Workload Measures

In this experiment, workload was measured objectively through response times to online instant message questions, and subjectively through a modified NASA TLX survey administered after each experimental mission scenario.

5.4.1. Secondary Workload

As expected, secondary workload significantly increased from low to high re-planning ($F(1,12) = 7.1$, $p=0.02$). However, secondary workload was not statistically significant across schedule management DST type. There was no significant interaction between the independent variables. Secondary workload significance across re-planning level validated the difference between the two mission scenarios in terms of operational tempo and the resulting workload subjects experienced. Non-significance of secondary workload across DST showed that the addition of either StarVis DST implementation to MAUVE did not statistically increase operator workload. This was a positive result, as the StarVis DSTs were supposed to assist operators in managing workload throughout the mission and not add to it.

5.4.2. Subjective Workload

Figure 5-3 shows a box plot of subjective workload scores for the different schedule management DSTs under the two re-planning levels. Subjective workload was marginally significant in both re-planning level ($F(1,12)=4.1$, $p=0.067$) and schedule management DST type ($F(2,12)=3.8$, $p=0.052$). There was no significant interaction between the independent variables. Tukey post hoc comparisons for subjective workload were only marginally significant between Local and Q-Global StarVis subjects ($p=0.053$). Local StarVis subjects generally reported lower subjective workload ratings than subjects using Q-Global StarVis. Statistically, subjective workload for Local StarVis subjects was no different across re-planning levels. Pearson correlations found that subjects with the best performance did not perceive workload to be high (r=-0.497, p=0.005). These results show that Local StarVis was successful in increasing performance while maintaining fairly steady and low perceived workload, even under different operational tempos.

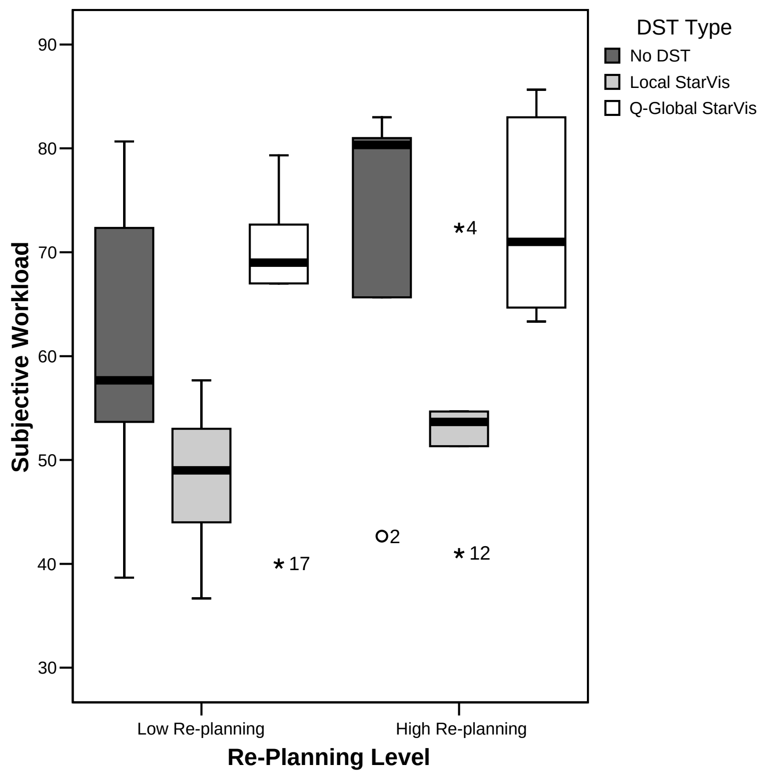


Figure 5-3: Subjective workload box plot.

5.4.3. Comparison of Workload Measures

Across schedule management DST, secondary and subjective workload results related very different stories about operator workload. Non-significance in secondary workload showed that there was no difference in spare capacity across subjects using different DSTs. Subjective workload results, however, showed that subjects using different DSTs had marginal differences in perceived workload. The incongruity between the two workload measures may be due to their different focuses within the multi-UAV mission. Secondary workload was measured through response times to online instant message questions, while subjective workload was calculated from the NASA TLX survey which queried subjects about overall mission elements. Because answering instant message questions was ranked as the lowest priority mission objective (see Appendix B), it is possible that all subjects, regardless of DST assignment, had statistically similar response times because all focused on higher mission priorities. Because the NASA TLX survey was more comprehensive in addressing different elements of workload across the whole mission, it appears to be more appropriate for assessing the mental workload of subjects in this experiment.

5.5. Situation Awareness

5.5.1. Situation Awareness Score

Subject situation awareness, as measured by the SA score (Section 4.4.2), significantly decreased from low to high re-planning level ($F(1,12)=45.9$, $p<0.001$), and was statistically significant across schedule management DST type ($F(2,12)=9.8$, $p=0.003$). There was a significant interaction between the independent factors ($F(2,12)=5.7$, $p=0.018$). As shown in Figure 5-4, this interaction was due to the relatively small difference in SA for all subjects under low re-planning, whereas under high re-planning, subjects using the Local StarVis had significantly higher SA scores than other subjects. Tukey post hoc comparisons established that subjects using the Q-Global StarVis had no difference in SA compared to those subjects without a schedule management DST ($p=0.775$), whereas subjects using the Local StarVis had superior SA compared to those subjects using both Q-Global StarVis ($p=0.013$) and no DST ($p=0.004$).

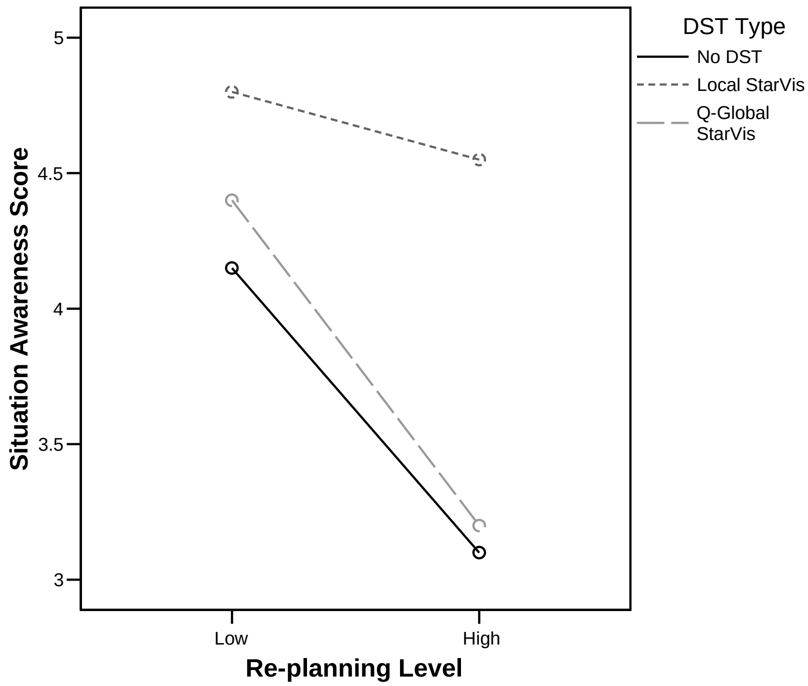


Figure 5-4: Situation awareness results

SA score results under high re-planning reveal that Local StarVis subjects had significantly better awareness of the mission situation, both in the present and future projective sense, than subjects in the other experimental conditions. Additionally, across both mission scenarios, Local StarVis produced no statistical difference in SA, even though workload level changed. They did not lose sight of present mission activities while managing a complex schedule under different amounts of workload. Subjects without a DST and those using Q-Global StarVis, however, had statistically the same SA scores across both levels of re-planning. Under high re-planning, these subjects' scores revealed severe decreases in current and projective awareness of the mission situation.

SA was also strongly correlated with performance ($r = 0.828$, $p < 0.001$). This was slightly expected as some, but not all, of the SA score indicators were similar to components of the performance score. The correlation indicates that subjects who performed mission objectives also had good awareness and understanding of the current and future mission situation.

5.5.2. Critical Firing Events

The number of critical firing events, defined as incorrectly firing upon targets, measured global situation awareness for all subjects in each schedule management DST across both mission scenarios. A Chi Square test was used to statistically analyze this metric's results, and although not significant across DSTs, the trend for critical firing events was in favor of Local StarVis. There were a total of two critical firing events by subjects using no DST and three by Q-Global StarVis subjects. Not one critical firing event occurred for any subject using Local StarVis. These results imply that Local StarVis subjects may have had better global SA than subjects using the Q-Global StarVis or those without a DST.

5.6. Schedule Problem Mitigation Scores

Schedule problem mitigation scores measured how many schedule problems, either late arrivals or TOT conflicts, subjects eliminated or created during the mission.

5.6.1. Late Arrival Mitigation Score

Late arrival mitigation score significantly decreased from low to high re-planning level ($F(1,12) = 5.5$, $p=0.037$), but was not statistically significant across DST type. There was no interaction between the independent variables. When controlling for the schedule management DST, late arrival mitigation score significantly correlated with performance ($r = 0.560$, $p = 0.002$), revealing that one way in which the experiment's best performers achieved higher performance scores was through mitigating late arrivals. When broken out across the different DSTs, the correlation between late arrival mitigation and performance had the following results:

- No DST: $r = 0.715$, $p = 0.020$
- Local StarVis DST: $r = 0.721$, $p = 0.019$
- Q-Global StarVis DST: $r = 0.102$, $p = 0.780$

Thus, in the no DST and Local StarVis conditions, subjects who mitigated more late arrivals tended to have better performance. This relationship did not hold for Q-Global StarVis subjects; the number of late arrivals they mitigated did not correlate to how well they performed the mission management task. A marginally significant Kruskal-Wallis test showed that subjects using the Local StarVis mitigated more late target arrivals than subjects from the other two conditions ($\chi 2(2) = 0.102$, $\alpha = 0.1$).

5.6.2. TOT Conflict Mitigation Score

TOT conflict mitigation score significantly increased from low to high re-planning level ($F(1,12) = 26.6$, $p < 0.001$), but was not statistically significant across schedule management DST. There was no interaction between the independent variables. When controlling for the different DSTs, TOT conflict mitigation negatively correlated with performance ($r = -0.372$, $p = 0.047$). Parsed out across the different DSTs, only Local StarVis TOT conflict mitigation scores marginally, yet negatively, correlated with performance ($r = -0.609$, $p = 0.062$), signifying Local StarVis subjects who had higher performance mitigated fewer TOT conflicts.

Further investigation revealed this relationship could have been driven by the difference in the number of TOT conflicts subjects saw in a mission. Subjects without a DST experienced significantly more, and thus mitigated more, TOT conflicts than either

Local or Q-Global subjects ($p<0.001$). However, there was no statistical difference between the two StarVis conditions for either the number of TOT conflicts generated or mitigated. The higher rate of TOT conflict generation and mitigation for subjects without a DST was likely due to the difficulty they had in understanding long-term future impacts of TOT delay requests. By requesting many TOT delays, subjects without a DST tended to generate more TOT conflicts than they fixed.

5.6.3. Discussion

Overall, schedule problem mitigation results showed that subjects who performed the best prioritized late arrivals as the most critical schedule problem and minimized the creation of TOT conflicts through schedule changes. This strategy, utilized by subjects using either StarVis DST, was exhibited more by Local subjects who tended to mitigate more late arrivals. Most Local StarVis subjects realized the importance of fixing late arrivals, as this schedule problem preordained that targets would be missed unless action was taken. TOT conflicts, on the other hand, did not guarantee missed targets, only that a predicted future high workload period *could* lead to missed targets.

Post-experiment feedback forms validate this claim. One StarVis subject commented, "The indication of 'Late Arrival' [on the StarVis display] was useful. [The] TOT conflict indicator does warn me to be alert, but it has no influence on my planning." Subject strategies were confirmed quantitatively from the correlations; subjects who concentrated on late arrival mitigation and minimized TOT conflict generation tended to perform better.

TOT conflict mitigation results also agreed with previous research revealing that subjects who fixated on eliminating potential conflicts tended to do worse [35]. When considering schedule problem mitigation scores with TOT delay request results, it appears that Local StarVis subjects appropriately allocated their few delay requests to fix late arrival problems, a strategy not generally used by other subjects.

While these results describe schedule management strategies used by high-performing Local StarVis subjects, they suggest potential design changes to the StarVis configural display itself. Because subjects performed better when they favored fixing late arrivals over TOT conflicts, perhaps TOT conflict information should either be eliminated from the StarVis DST, be made less salient, or be offloaded to automation. A study following the StarVis experiment examined how a MAUVE-embedded bar-graph display showing only current and "what if" late arrival information, called BarVis, helped operators manage their mission schedule and achieve mission objectives [49]. Results from an experiment identical to the StarVis experiment found that subjects using BarVis did not perform better or have lower subjective workload than Local StarVis subjects [50]. Because this follow-on study manipulated both the visual appearance and information content of the configural DST, it was unable to conclude the exact cause of these results. Future studies should examine if BarVis' lower performance and higher perceived workload was due to the display's appearance or to its lack of TOT conflict information.

5.7. DST Interaction

Interaction with the StarVis "what if" mode, as measured by the number of times subjects selected or de-selected target checkboxes, was not significant across re-planning level or StarVis DST type, and there was no significant interaction between the independent variables. There was a marginal negative correlation between DST interaction and performance ($r=-0.391$, $p=0.088$). However, when controlling for StarVis DST type, DST interaction and performance did not correlate. Thus, more experimental data is needed to determine if StarVis "what if" mode use detracted from performance.

5.8. Post-Experiment Subjective Responses

General feedback was obtained from subjects upon completion of the experiment through a post-experiment feedback questionnaire (Appendix F). A sample of responses pertaining to the different schedule management DSTs are presented and discussed here.

5.8.1. Timeline Only (No DST) Subject Responses

All subjects who were provided with only the color-coded mission timeline (no DST) found it useful, even crucial, for schedule management. However, a few subjects wanted a timeline that showed the mission plan farther than fifteen minutes in into the future so they could better understand how early schedule changes affected the mission schedule later on. Subjects without a DST also felt time pressure in making decisions. For example, one subject wished for “a longer planning horizon.”

Subjects using only the timeline pointed out the very drawbacks that prompted development of the StarVis DSTs. One subject noted that the timeline did not suggest alternate decisions for schedule management, and that schedule re-planning was virtually trial and error. Another subject commented that the timeline did not show the effects of schedule management decisions until after TOT delay requests were granted. Additionally, a subject stated “Sometimes I expected TOT conflicts when there were none,” which brings into question the saliency of schedule problems on the timeline. One subject also appears to have disregarded the probability bar associated with the TOT delay request button and instead fixated on obtaining TOT delays, a behavior exhibited from a previous study [35]: “The low probability of getting a TOT delay discouraged me first from requesting delays. Later on I understood that I simply had to insist more to get the delay.”

5.8.2. Local StarVis DST Subject Responses

Many subjects who used the Local StarVis found it confusing and frustrating, and some did not like or trust the DST. One subject tried to double check "what if" predictions because the display seemed unreliable and the subject thought it was erroneously predicting future TOT conflicts. The best performer in the entire experiment claimed that he did not use StarVis much at all. Instead, he claimed to have used the timeline plus mental projections to predict the consequences of schedule management decisions.

One subject found the Local StarVis DST "fairly helpful in making decisions" but thought that it did not consider more complex schedule decision scenarios, such as what would happen to the schedule if multiple TOTs were changed. This subject commented that the Local StarVis did not provide any recommendations on optimal courses of action for fixing the overall schedule, but it did help identify current schedule problems and was useful "in analyzing the effectiveness of basic solution possibilities."

One subject was confused by the graph nature of the StarVis configural display, stating that he "kept trying to interpret it as a graph, but actually nothing was being plotted." Another subject expressed that StarVis did not preview or project the change(s) to an individual UAV's schedule before a target was added to or removed from its timeline. Generally, most Local StarVis subjects praised the timeline, stating they found it more useful for making schedule changes, and criticized the Local StarVis, particularly its "what if" mode, saying that they rarely used or needed it.

5.8.3. Q-Global StarVis DST Subject Responses

Some of the responses from Q-Global StarVis subjects mirrored those from Local StarVis users. Q-Global StarVis subjects also found the StarVis display to be confusing and difficult to use, particularly when in "what if" mode. One subject expressed that it was hard to match the StarVis triangles to the targets with problems, and commented that the Q-Global StarVis was helpful only when there were few schedule problems and only a few delays would be necessary to resolve them. Another subject commented that selecting multiple checkboxes on the Q-Global design was too time consuming and did not indicate the best course of action for requesting TOT delays. Some subjects felt that too much information was shown in too small a space when the Q-Global StarVis was in "what if" mode. This effect could have increased examination time of the Q-Global DST and hindered quick judgment. Other subjects expressed the desire for "what if" schedule information to be overlaid on the timeline, or that while in "what if" mode, the gray triangles on the StarVis would disappear to indicate potentially fixed current schedule problems.

More than one subject admitted that the Q-Global StarVis was useful in indicating late arrivals, and one even stated that it "provided a quick and intuitive overview of the situation" once they got used to its format. However, two subjects did not find the TOT conflict indications to be useful, and said that they did not consider TOT conflicts in mission schedule re-planning. All Q-Global StarVis subjects liked the DST, believed it was okay, or thought it neither helped nor hindered their ability to manage the multi-UAV mission. Subjects preferred using it to understand current schedule problems, but disliked its projective "what if" mode.

5.9. Summary of Experimental Findings

Table 5-1 summarizes the statistical results for all dependent variables. The main experimental finding was that Local StarVis subjects achieved higher performance scores, generally reported lower subjective workload, and had higher situation awareness than subjects using the Q-Global StarVis or no DST. When managing the mission schedule, subjects using Local StarVis demonstrated an effective strategy by requesting TOT delays sparingly to mitigate late target arrivals, while simultaneously minimizing the number of created TOT conflicts.

Dependent Variable	DST Type	Level of Re-planning
Performance Score	0.003	<0.001
Number of TOT Delay Requests	0.32	0.065
Percentage of Approved TOT Delay Requests	0.53	0.73
Secondary Workload	0.63	0.02
Subjective Workload	0.052	0.067
SA Score	0.003	<0.001
Late Arrival Mitigation Score	0.22	0.037
TOT Conflict Mitigation Score	0.74	<0.001
DST Interaction	0.29	0.24

Table 5-1: Statistical summary of experimental results (p-values).

The small number of experimental subjects may have contributed to marginally significant or non-significant results across some of the dependent measures, such as the number of TOT delay requests, subjective workload, and DST interaction. It is possible that with more subjects, results could show stronger statistical similarities or differences in dependent variables across the different DSTs. To reach more definitive conclusions, a larger number of subjects are needed.

The disparity in experimental measures between the Local and Q-Global StarVis DSTs was surprising as both utilized the same configural display but employed different "what if" mode implementations. As discussed in Chapter 3, both StarVis DSTs identically depicted current schedule problem information. The only difference between the DSTs was in the operation of their “what if” modes. Thus, it appears that the differences in “what if” modes between the two StarVis DSTs contributed to the inequality in the experimental results. However, as shown by the DST interaction results, only a marginal negative correlation existed between interaction with the “what if” mode and performance. It was decided that additional retrospective analysis was needed to examine how the Local and Q-Global StarVis DSTs caused subjects to interact differently with their respective “what if” modes. This retrospective analysis is discussed in the next chapter.

6. Retrospective Analysis

This chapter describes retrospective analysis conducted to further explain the experimental results described in Chapter 5.

6.1. Hypotheses

A variety of factors could have led to the observed differences in experimental metrics between the Local and Q-Global StarVis DSTs, including the number of times subjects used the StarVis "what if" mode (called projection), the total length of time the "what if" mode was active during a mission, the portion of a mission during which subjects actively used the "what if" mode, how much total "what if" information subjects saw in a mission, how much information subjects saw per each projection, and the change in displayed information quantity when transitioning from current problems mode to "what if" mode.

It was hypothesized that the performance difference between the two StarVis DSTs was primarily due to the Q-Global design overloading subjects with information when they used its "what if" mode. In the Q-Global design, selecting a target checkbox to engage the "what if" mode often caused many yellow and split gray-yellow triangles to appear across multiple StarVis displays. This large volume of information may have increased perceptual and cognitive processing of Q-Global subjects. Because Q-Global subjects had to examine all the StarVis displays and synthesize the provided information, they might not have easily understood the potential effects of delaying selected target(s), which may have degraded their performance and increased their perceived workload.

In contrast, selecting one target checkbox in the Local StarVis design only affected one UAV's StarVis display. When using its "what if" mode, Local StarVis subjects only needed to look at displays corresponding to the checkboxes they selected. Thus, Local subjects had less information to perceive and analyze when using the "what if" mode, even if they selected multiple checkboxes or scanned StarVis displays with unselected checkboxes.

Although it was suspected that information overload caused the differences in experimental measures between Local and Q-Global StarVis subjects, retrospective analysis was conducted to analyze how different factors pertaining to "what if" mode interaction may have caused the differences. The methods, results, and discussion of this additional analysis compose the remainder of this chapter.

6.2. Analysis Overview

For the retrospective analysis, only the high re-planning mission scenario was examined, because as experimental results showed, differences in metrics between subjects using different DSTs were more extreme under higher workload. For example, SA results showed that all subjects had statistically identical SA scores under low re-planning, but under high re-planning, Local StarVis subjects maintained higher SA than those subjects who used the Q-Global StarVis or no DST.

Only subjects who actually used the StarVis "what if" mode were analyzed retrospectively, as the goal of the additional analysis was to pinpoint the differences in factors pertaining to "what if" usage. In the high re-planning mission, all five Local and four out of five Q-Global StarVis subjects used the "what if" mode at some point.

6.3. Analyzed Factors and Data Collection

Figure 6-1 illustrates factors relating to "what if" mode interaction, which could explain the differences between Local and Q-Global StarVis subjects in performance and other experimental metrics. Data pertaining to these factors was collected through examination of experiment data files and Camtasia® screen capture video recordings.

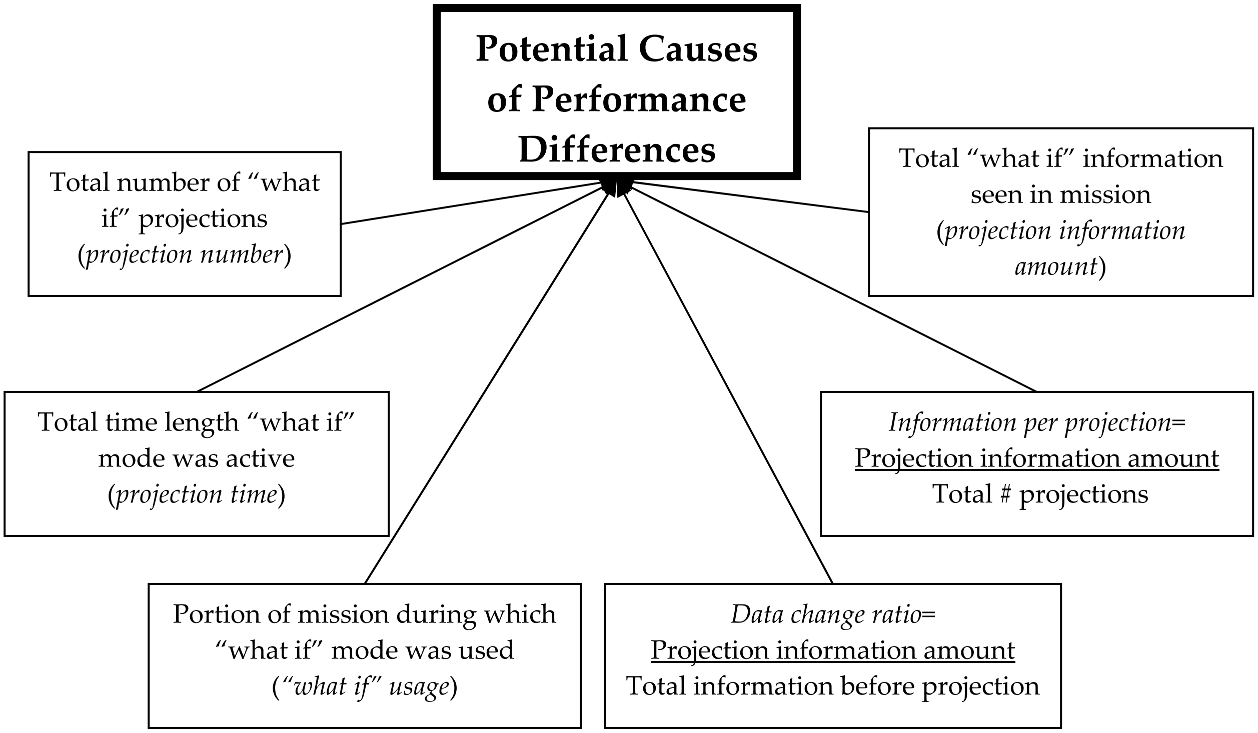


Figure 6-1: Potential "what if" mode interaction factors influencing differences in performance between Local and Q-Global StarVis DST subjects.

The total number of "what if" projections (henceforth called *projection number*) was measured for each subject by counting the number of times in the mission they engaged the StarVis "what if" mode by selecting a target checkbox. Multiple projections on the same target were included in this count, even if a subject quickly selected or de-selected ("toggled") the same target checkbox. This factor was examined to see if activating the StarVis "what if" mode affected performance.

The length of time the "what if" mode was active for the mission (henceforth called *projection time*) was measured for each subject by summing the amount of time checkboxes were selected on the StarVis display. Checkboxes could only be selected by subjects, but could either be cleared by subjects or by changes in current schedule problems, such as when TOT delay requests were approved. Projection time was analyzed to see if the length of time the "what if" mode was active for the mission caused subjects to overly focus on schedule management, thus affecting their performance. In contrast to projection time, the portion of the mission during which subjects used the "what if" mode (henceforth called *"what if" usage*) represented the length of mission time when subjects used the StarVis projective capability. "What if" usage was measured from the first projection in the mission to the last StarVis clearing. This factor was investigated to see if greater "what if" mode usage throughout an entire mission influenced subject performance. Both projection time and "what if" usage factors were measured in seconds.

To examine how much information each StarVis DST provided, information quantities were defined by bit counts. Bit values for triangles were defined in the following manner: gray and yellow triangles on the StarVis display each represented one bit, while split gray-yellow triangles represented two bits. Split triangles represented two bits because they predicted two pieces of information: a current schedule problem could continue to exist even after an approved TOT delay request. For each subject, bit counts were collected for the total amount of information shown across the StarVis displays before "what if" mode activation (henceforth called *pre-projection information amount*), and the total amount of information shown when the "what if" mode was engaged (henceforth called *projection information amount*). For data collection purposes, projection information was parsed into counts of grey, yellow, and split triangles. Pre-projection and projection information amounts were measured to examine if some StarVis subjects saw more schedule management decision support information than others, which could have caused information overload and hindered decision-making.

The ratio of projection information amount over projection number (henceforth called *information per projection*) represented the average amount of "what if" information (in bits) subjects saw each time they projected. Information per projection was measured in bits per projection and was examined to determine if the Q-Global StarVis "what if" mode showed subjects more information each time it was used than the Local "what if" mode. Discrepancies between the two StarVis DSTs in how much "what if" information per projection was shown could explain the differences found in performance and perceived workload. The ratio of projection information amount over pre-projection information amount (henceforth called *data change ratio*) measured the change in data subjects saw before and after engaging StarVis' "what if" mode. This ratio represented a dimensionless quantity, with values less than 1 indicating a decrease in information when the "what if" mode was used, and values greater than 1 signifying information increase. Large changes in information quantity between transitions from the StarVis current problems mode to "what if" mode could have made it difficult for operators to understand how their current schedule compared to their "what if" schedule in terms of number and types of schedule problems. Large information changes could have complicated schedule management decision-making.

6.4. Retrospective Analysis Results

The following subsections provide results and discussion of how various StarVis "what if" mode interaction factors may have affected subject performance. For analysis, descriptive statistics and box plots were used to determine data trends. Correlations were also drawn between each of the factors and performance, $\alpha = 0.05$.

6.4.1. Performance and Projection Number

On average across the high re-planning mission, Q-Global StarVis subjects used the "what if" mode twice as much as Local subjects (Local mean = 6.4, Local standard deviation = 4.7; Q-Global mean = 12.8, Q-Global standard deviation = 10.9). Nonetheless, of the subjects who used the StarVis "what if" mode under high re-planning, Local

StarVis subjects performed almost two times better than Q-Global subjects (Local mean = 600.5, Local standard deviation = 73.3; Q-Global mean = 313.1, Q-Global standard deviation = 148.0. Despite this trend, under high re-planning, performance and projection number were not correlated.

6.4.2. Projection Time and "What If" Usage

On average, projection time was longer for Local StarVis subjects than for Q-Global subjects (Local mean = 95.8 sec., Local standard deviation = 90.0 sec.; Q-Global mean = 60.3 sec., Q-Global standard deviation = 71.5 sec.). However, "what if" usage was, on average, longer for Q-Global than Local StarVis subjects (Local mean = 376.6 sec., Local standard deviation = 447.7 sec.; Q-Global mean = 568.3 sec., Q-Global standard deviation = 287.0 sec.). These combined results show that although Q-Global subjects used the StarVis "what if" mode to a greater degree throughout the experiment, the projective mode was less active during the mission. When considering projection number, these results verify a Q-Global subject behavior observed in screen capture videos: "what if" mode toggling. When using the "what if" mode, Q-Global StarVis subjects tended to quickly select and deselect the same target checkbox, causing short second-long durations when the mode was active. Toggling explains how Q-Global StarVis subjects, who on average tended to project more, had shorter average projection times than subjects using Local StarVis.

Of particular interest was the behavior of a Q-Global StarVis subject who performed the worst under high re-planning. This subject not only performed single-target toggling, but also showed a unique toggling behavior, called list-toggling, where he would quickly select and deselect, in order, all targets in the schedule problem list. Subjects using the Local StarVis rarely toggled single targets and never list-toggled.

Although these results verify that toggling was common with Q-Global StarVis subjects, neither projection time nor "what if" usage correlated with performance.

6.4.3. Pre-Projection and Projection Information Amount

Differences in pre-projection information amount between the different StarVis DSTs are shown in Figure 6-2. On average, pre-projection information amount was three times higher for Q-Global than Local StarVis subjects (Local mean = 23.0 bits, Local standard deviation = 19.4 bits; Q-Global mean = 69.3 bits, Q-Global standard deviation = 72.6 bits). Because both StarVis DSTs identically displayed current schedule problems, this result indicates that Q-Global StarVis subjects tended to use their "what if" mode when their mission had many current schedule problems. In contrast, when using their "what if" mode, Local StarVis subjects had fewer current schedule problems than Q-Global subjects. However, there was no significant correlation between pre-projection information amount and performance under high re-planning.

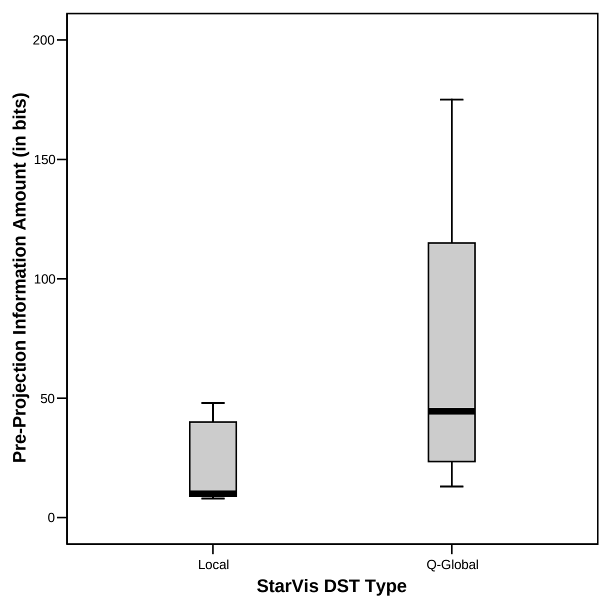


Figure 6-2: Pre-projection information amount box plot for both StarVis DSTs under high re-planning.

Projection information amount results show that on average in the high re-planning mission, Q-Global StarVis subjects saw four times more "what if" information than Local subjects (Local mean = 30.2 bits, Local standard deviation = 27.8 bits; Q-Global mean = 130.3 bits, Q-Global standard deviation = 142.3). This result was primarily driven by the number of split triangles seen by Local versus Q-Global StarVis subjects when using the "what if" mode. On average, Q-Global subjects saw over twelve times more split triangles than subjects using Local StarVis (Local mean = 4.2, Local standard deviation = 5.8; Q-Global mean = 54, Q-Global standard deviation = 63.3). When comparing the number of "what if" grey and yellow triangles, on average, Q-Global and Local StarVis subjects saw about the same number of grey triangles when using the "what if" mode (Local mean = 18.2, Local standard deviation = 12.8; Q-Global mean = 15.8, Q-Global standard deviation = 11.7), and Q-Global subjects saw almost two times more yellow triangles than Local StarVis subjects (Local mean = 3.6, Local standard deviation = 3.8; Q-Global mean =6.5, Q-Global Standard Deviation = 4.4).

When in "what if" mode, split triangles dominated the Q-Global StarVis, while grey triangles were primarily present on Local StarVis. The latter result was expected, as Local StarVis only showed "what if" information across configural displays corresponding to selected checkboxes. Often only one Local StarVis display projected "what if" information, while the other displays continued to show current schedule problems through grey triangles. Because selecting a checkbox in the Q-Global StarVis projected potential TOT delay effects across all the configural displays, many split triangles emerged because delaying the selected target would unlikely fix unrelated schedule problems.

The sheer number of split triangles shown on the Q-Global StarVis during "what if" mode was not expected. This large average may have been caused by the tendency of Q-Global subjects to use the "what if" mode when there were many current schedule problems, as seen by pre-projection information amount results. The large number of split triangles may have made it difficult for Q-Global subjects to compare the consequences of delaying one target on the overall schedule versus taking no action. Figure 6-3a-b shows box plots for the number of (a) grey and yellow, and (b) split triangles Local and Q-Global StarVis subjects saw when using their respective "what if" modes. Figure 6-3c shows a box plot representing the total projection information amount, in bits, seen by both StarVis conditions.

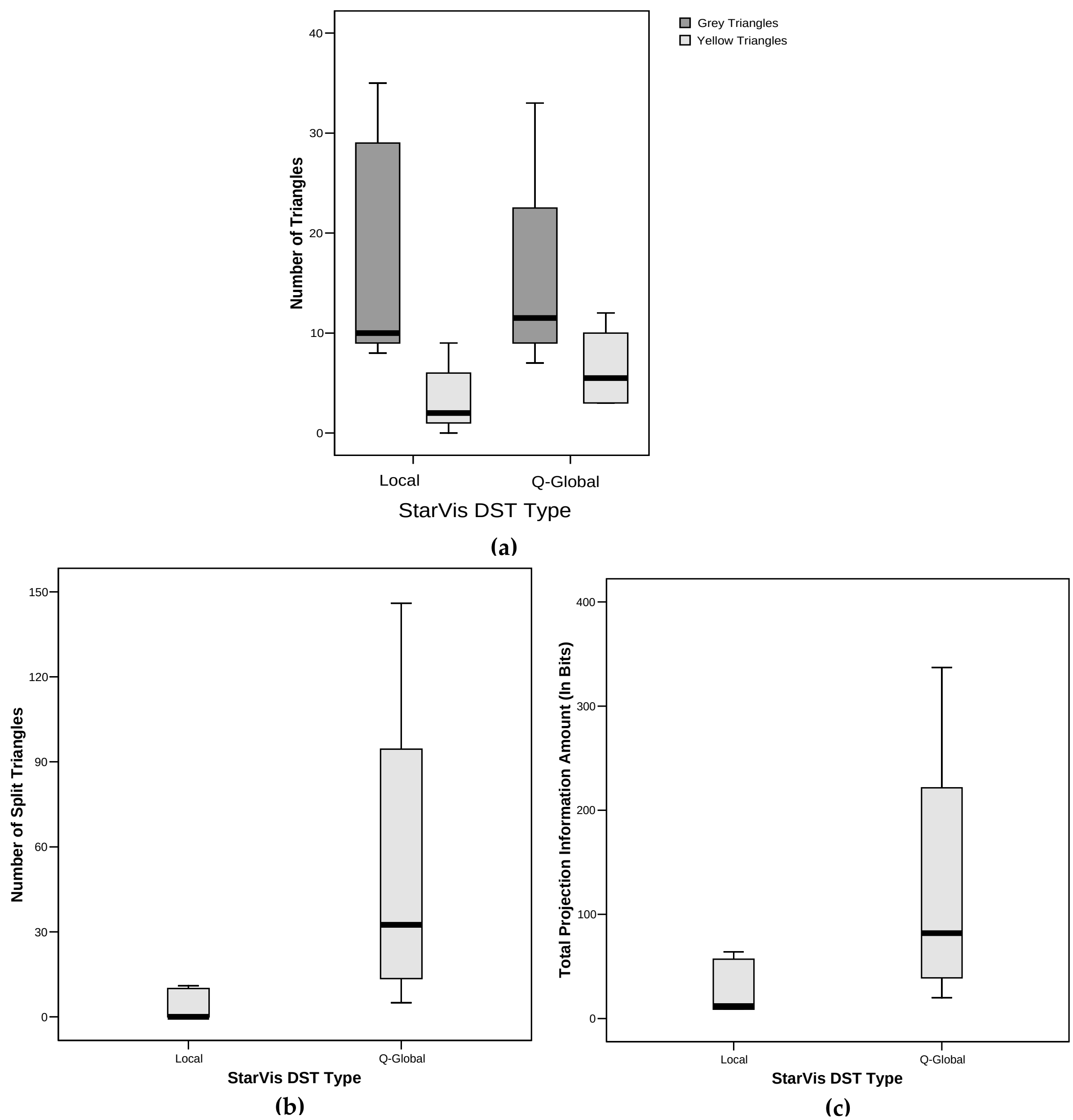


Figure 6-3: Box plots of information seen by subjects during StarVis "what if" mode. (a) Number of grey and yellow triangles (b) Number of split triangles (c) Total projection information amount.

6.4.4. Information per Projection

Information per projection was calculated to determine the average information amount (in bits) subjects saw per each "what if" mode use, thus controlling for differences in "what if" usage between subjects. This factor was examined to see if differences in "what if" information quantities (per projection) contributed to the suspected information overload that Q-Global StarVis subjects may have experienced. Figure 6-4 shows a box plot for information per projection. On average, Local StarVis subjects saw approximately half the amount of information per projection than subjects using the Q-Global StarVis (Local mean=5.0 bits/projection, Local standard deviation = 2.8 bits/projection; Q-Global mean=8.7 bits/projection, Q-Global standard deviation = 2.9 bits/projection). Thus, Q-Global StarVis subjects saw more information per each projection than Local Subjects, which could have contributed to their potential information overload.

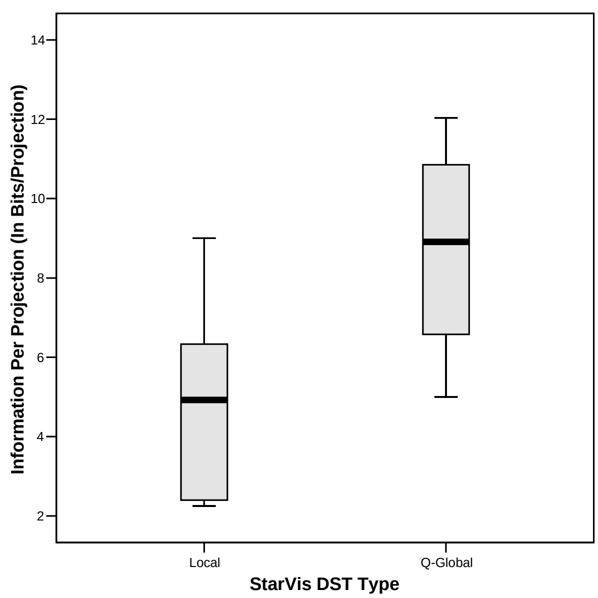


Figure 6-4: Box plot for information per projection across the high re-planning mission.

A marginal Pearson correlation between high re-planning performance and information per projection found that StarVis subjects performed better when they saw less “what if” information per projection ($r=-0.658$, $p=0.054$). This mirrors the descriptive statistics, as Local StarVis subjects saw less information per projection and performed better than Q-Global subjects. The lower performance exhibited by Q-Global subjects may have been caused by the potential information overload spawned by the Q-Global “what if” mode.

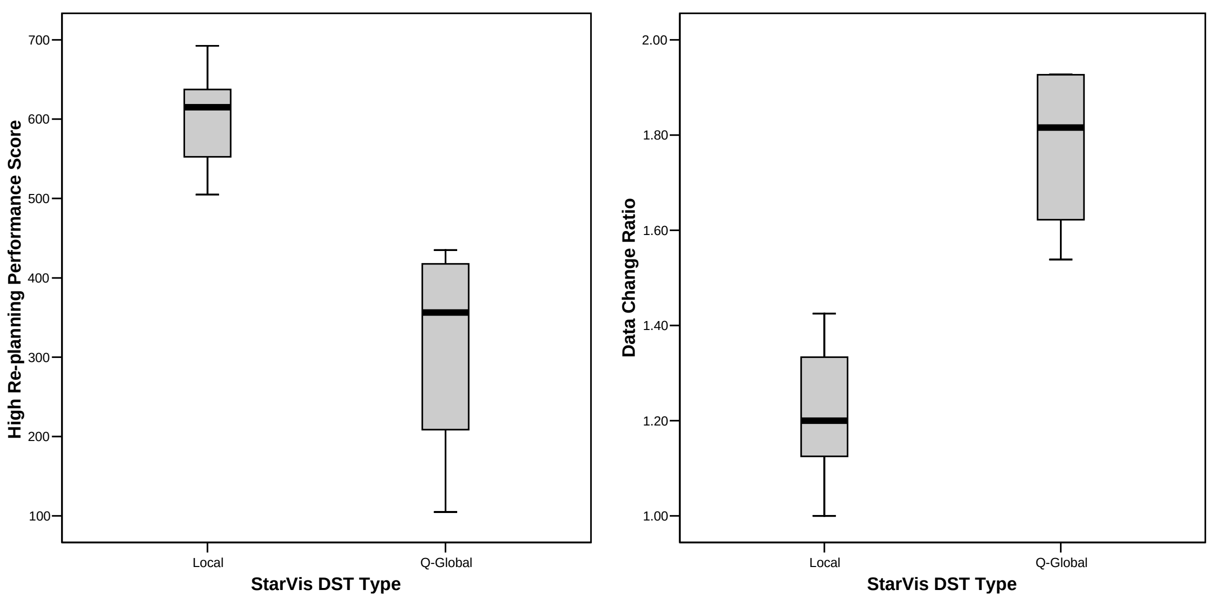
6.4.5. Data Change Ratio

Data change ratio was measured to determine if subjects using different StarVis designs saw different quantities of information change between transitions from current problems mode to “what if” mode. Data change ratio differed from information per projection as it did not examine how much information subjects saw each time they used the StarVis “what if” mode. Instead, data change ratio studied how information amounts increased or decreased when the StarVis transitioned from current problems mode to “what if” mode. This factor was studied to determine if large changes in information with StarVis mode transition affected operator performance.

On average in the high re-planning mission, Local StarVis subjects saw 1.2 times more data when the “what if” mode was engaged than immediately before (Local standard deviation = 0.17), while Q-Global StarVis users saw 1.8 times more data (Q-Global standard deviation = 0.19). When examining individual subject data change ratios, all Local subjects had data change ratios less than 1.5, while all Q-Global subjects who used the “what if” mode had data change ratios greater than 1.5. Under high re-planning, the worst performing Local subject had the highest data change ratio (1.43) out of all Local StarVis subjects. This observation mirrored the strong negative correlation between high re-planning performance and data change ratio ($r = -0.766$, $p =$

0.016). Figure 6-5 shows this correlation graphically through box plots for high re-planning (a) performance score and (b) data change ratio.

Based upon the strong correlation with performance, data change ratio could help explain why Local StarVis subjects outperformed Q-Global subjects in the multi-UAV mission supervision task. Because the Q-Global StarVis showed much more "what if" information than the Local design, the larger change in data when transitioning to "what if" mode may have caused Q-Global subjects difficulty in understanding the difference between current and "what if" schedules. This may have prompted Q-Global subjects to toggle the "what if" mode in order to see its projective information "flash" out at them. Excessive toggling could have been costly in terms of time and cognitive capacity when subjects fixated upon selecting and de-selecting the "what if" mode. Local StarVis subjects may not have toggled because they saw much less information change when they transitioned from current problems mode to "what if" mode.



(a) **(b)**

Figure 6-5: High re-planning mission box plots for (a) performance score and (b) data change ratio

6.4.6. Other Interesting Findings

Examination of individual subject data revealed other intriguing findings. Under high re-planning, the top three Local StarVis performers used the "what if" mode the least. They were also the only StarVis subjects to earn high re-planning performance scores over 600 points (with a maximum score of 692.5), had the lowest total "what if" bits (with no split triangles), and had the smallest data change ratios. Thus, the best performing Local StarVis subjects tended to use the "what if" mode sparingly and only when they had few current schedule problems.

The Q-Global subject who used the "what if" mode the least (and not at all under high re-planning) out of all subjects was the best performing Q-Global subject overall. In contrast, the overall worst-performing Q-Global subject used the "what if" mode the most out of all subjects across the entire experiment. These results potentially signify that both StarVis designs were most useful in summarizing current schedule problems, but could hurt performance if subjects used the "what if" mode.

This speculation mirrors comments from post-experiment feedback forms, as well as investigator observations of subjects during experimental training. Although Local StarVis subjects outperformed all others, they did not necessarily find the Local "what if" mode useful. The Local StarVis subject who performed the best out of all subjects commented "I did not use the [Local StarVis] much at all, I did the same things it did in my head." During training, most StarVis subjects, regardless of implementation, found it easy to understand how the configural display represented current schedule problems, but both "what if" mode designs confused subjects and often required explanation beyond provided training documents.

Examination of Camtasia® video recordings found that neither Local nor Q-Global StarVis subjects used their "what if" mode designs as had been intended. Under high re-planning, only two Local subjects simultaneously engaged the "what if" mode on multiple StarVis displays. Because the Local StarVis was designed to concurrently show subjects the potential effects of multiple solution possibilities for one schedule problem, most Local subjects did not use the "what if" mode as anticipated. Instead, Local StarVis subjects tended to assess potential schedule problem solutions individually, even when addressing TOT conflicts.

Q-Global StarVis subjects also did not use their "what if" mode as had been anticipated. Under high re-planning, only one Q-Global StarVis subject simultaneously engaged multiple checkboxes while using the "what if" mode. This subject, who performed the worst out of the Q-Global group, activated a multi-target "what if" five times in the high re-planning mission. Multi-target projections could have further confused the subject because of excessive "what if" information, which could have contributed to his degraded performance.

6.5. Summary of Retrospective Analysis

The results of the retrospective analysis, summarized in Table 6-1, explain why performance differed between Local and Q-Global StarVis subjects. Retrospective analysis suggested that subjects using the Q-Global StarVis likely performed worse because they experienced information overload when using their DST's "what if" mode. When using this projective mode, Q-Global subjects saw more information per projection and a larger change in information with transition from current problems mode to "what if" mode. This effect may have been exacerbated because Q-Global subjects tended to use the "what if" mode when they had many schedule problems. They also used the projective mode for greater portions of the experiment. The possible confusion caused by the Q-Global StarVis "what if" mode may have played a role in Q-Global subjects' lack of appropriate attention to late target arrivals. Perhaps because of the distracting and confusing nature of the Q-Global StarVis, subjects had difficulty realizing that late target arrivals were more critical than TOT conflicts.

Generally, the best performers across both StarVis designs tended to use the "what if" capability sparingly. Experimental results show that the addition of a StarVis configural display in a Local implementation improved performance. However, this could be due more to the display's representation of current schedule problems than by its "what if" capability. This speculation is supported by retrospective analysis results which found that Q-Global StarVis subjects who used the "what if" mode sparingly tended to have higher performance. However, future studies are needed to definitively determine if the StarVis current problems mode played a greater role in subject performance than the "what if" mode.

While it is possible that the StarVis was most useful for alerting subjects of current schedule problems, it is also possible that "what if" information provided by the Local StarVis was enough to help its subjects make "good enough" schedule management decisions. Although the Local StarVis DST did not provide global information about the potential effects of schedule changes across all UAVs (as the Q-Global StarVis did), it also did not overwhelm subjects with excessive information.

Factor	Local Mean	Q-Global Mean	$\frac{\text{Q-Global Mean}}{\text{Local Mean}}$	Correlation w/ Performance
Projection Number	6.4	12.8	2.0	p=-0.36, r=0.34
Projection Time (sec)	95.8	60.3	0.6	p=0.08, r=0.84
"What If" Usage (sec)	376.6	568.3	1.5	p=-0.23, r=0.55
Pre-Projection Information Amount (bits)	23.0	69.3	3.0	p=-0.44, r=0.24
Projection Information Amount (bits)	30.2	130.3	4.3	p=-0.45, r=0.23
"What If" Split Triangle Count	4.2	54	12.9	p=-0.47, r=0.20
Information Per Projection (bits/projection)	5.0	8.7	1.7	**p=-0.66, r=0.054**
Data Change Ratio	1.2	1.8	1.5	**p=-0.77, r=0.016**

Table 6-1: Summary of key retrospective analysis results.

7. Conclusion

This chapter summarizes the motivation and objectives for this research, presents the key findings, and makes recommendations for future work in multi-UAV mission schedule management research.

7.1. Study Motivation and Research Objectives

As UAVs become more common in civilian and military applications, increasing levels of autonomy will lead to one operator supervising multiple UAVs. Under this operational scheme, critical human factors issues of performance, mental workload, and situation awareness are of concern. Time-critical multi-UAV missions involving uncertainty or potential emergent events will likely require significant operator knowledge-based reasoning to manage the complex mission, and in particular, its schedule. Multi-UAV operators will need to manage the schedule in real time, predict potential high workload periods and other schedule problems, and mitigate them in advance. In making schedule alterations, operators should understand beforehand the effects of their actions in the face of schedule uncertainty. A decision support tool for schedule management would likely help operators to understand potential schedule issues and the possible consequences of actions taken to address them.

The main contribution of the research presented in this thesis was the development and evaluation of a graphical decision support tool for schedule management of multiple UAVs in a time-critical mission. A configural display called StarVis was designed to provide information about current schedule problems as well as the possible long-term consequences of potential schedule changes to address those problems. StarVis was implemented into the MAUVE simulation as two different DST designs, Local StarVis and Q-Global StarVis, which differed in how they presented projective "what if" information. A human-subject experiment evaluated the different graphical StarVis DSTs in a multi-UAV time-critical targeting mission by measuring human performance, workload, and situation awareness.

7.2. Findings

An experiment comparing the two StarVis DSTs and a no DST control condition found that subjects using the Local design best achieved mission objectives as defined by a performance score. Local StarVis subjects also generally experienced lower perceived workload and had higher situation awareness than subjects using no DST or the Q-Global StarVis. When addressing schedule problems, Local StarVis subjects generally attended to fixing late target arrivals instead of mitigating potential high workload periods (TOT conflicts), a strategy not used by Q-Global subjects. Subjects that had no DST or used the Q-Global StarVis had statistically no difference in performance, subjective workload, and situation awareness in the experiment.

Performance differences between the two StarVis DSTs merited additional retrospective analysis. This analysis found that Q-Global StarVis subjects, when using their DST's "what if" mode, were likely overloaded with information because of the mode's projection of potential future schedule states across all of the StarVis configural displays. In contrast, Local StarVis only projected potential future schedule states across one or two StarVis configural displays. Under high re-planning, Q-Global StarVis subjects saw, on average, almost twice the amount of information per projection than Local StarVis subjects. On average, they also saw almost two times more information after engaging the "what if" mode than immediately before. Because Q-Global StarVis subjects had difficulty in understanding information differences between the DST's two modes, they tended to exhibit toggling behavior. The Q-Global StarVis "what if" mode provided a large volume of information which subjects were unable to quickly synthesize in order to make informed schedule management decisions.

Experimental trends showed that the best performers generally used the "what if" mode sparingly, a mannerism exhibited more by Local subjects. Thus, it is possible that the particular strength of the StarVis configural display was in its unified presentation of current schedule problems, instead of in its "what if" capability. It is possible, but not proven, that Local StarVis subjects may have performed better because they used their "what if" mode less than Q-Global subjects.

7.3. Recommendations and Future Work

Though the results of this thesis indicate that the Local StarVis configural DST improves multi-UAV supervisor performance, perceived workload level, and situation awareness by providing schedule management assistance, further investigations are warranted. The following are recommendations for future follow-on work based upon the research presented in this thesis.

- The results from this experiment called into question the severity of TOT conflicts in affecting achievement of mission objectives. Because TOT conflicts did not seem to be the most critical schedule problem, future research should examine how TOT conflict mitigation could be off-loaded to automation to identify, analyze, and solve potential high workload periods in advance, possibly with human operator involvement.

- The distinctiveness of the StarVis configural display could have influenced how quickly subjects discerned its represented information. Future studies should explore if a different configural display shape (similar to BarVis [49]) showing the same information as Local StarVis improves or degrades human performance, workload, and situation awareness.

- Experimental trends challenged whether a schedule management "what if" tool actually helped operators manage the mission schedule and achieve mission objectives. A StarVis configural display without "what if" capability should be tested against the Local StarVis DST described in this thesis to see whether Local StarVis subjects actually benefit from having projective decision support.

- While the StarVis DSTs provided multi-UAV operators with information to make schedule management decisions, they did not offer recommendations on how to locally or globally mitigate schedule problems and/or manage the overall mission schedule. Partially or fully automated schedule management recommendations should be developed and tested, potentially with schedule management decision support visualizations or tools.

- Both the Local and Q-Global StarVis DSTs parsed out schedule problem information for each UAV. A "Global StarVis" representing all mission schedule problems on one configural display should be developed and tested, as it could better assist operators in understanding the consequences of schedule management decisions across the entire mission.

- Further investigation should study toggling of display affordances and why subjects may exhibit this strategy. Inquiry into toggling strategies may influence future display affordance design.

Appendix A: Descriptive Statistics

This appendix outlines the participant pool demographics, as well as provides descriptive statistics on the dependent measures.

A.1 Demographics

Category	N	Min.	Max.	Mean	Std. Dev.
Age (years)	15	20	31	24	3.30
RPV Experience (hours)	3	40	70	53.33	15.28
Pilot/Flight Experience (hours)	6	1	200	93.5	83.18
Student (Y/N)	13Y, 2N	-	-	-	-
Gender (M/F)	11M, 4F	-	-	-	-

Table A-1: Descriptive statistics for experimental subject demographics.

A.2 Dependent variables

Category	N	Min.	Max.	Mean	Std. Dev.
Performance Score	30	105	850	547.92	186.01
Number of TOT Delay Requests	30	2	51	17.73	11.57
Percentage Approved TOT Delay Requests	30	20.0	77.8	50.81	16.53
Secondary Workload (seconds)	30	16	83	39.40	18.51
Subjective Workload	30	36.67	85.67	62.16	15.19
Situation Awareness	30	3	5	4.03	0.79
Late Arrival Mitigation Score	30	-7	0	-1.93	1.51
TOT Conflict Mitigation Score	30	-7	0	-3.10	1.54
DST Interaction	20	0	51	13.45	13.403
Number of Critical Firing Events	3	0	3	1.67	1.53

Table A-2: Descriptive statistics for all dependent variables

Appendix B: Supplemental Experiment Screens

The following screens were used in support of the StarVis-MAUVE experiments. They were displayed on the two screens not being used by the MAUVE 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 experimental mission objectives in priority order shown to subjects during the entire experiment.

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: Color coding reference table for UAV actions shown to subjects during training only.

Appendix C: Performance Score

The performance score was designed to measure the overall performance of a subject during an experimental session, and was adapted from [45]. A subject's score was based upon how well they achieved the mission objectives (Appendix B), with greater emphasis placed on higher priority objectives, while penalties were levied for actions that had negative consequences for the mission. Each completed mission objective or penalty occurrence had a specific amount of points associated with it, positive or negative, from which performance score was calculated. Each mission scenario had an associated performance score; thus each subject had two performance scores for the two completed mission scenarios.

C.1 Earned Points

An operator earned points toward their performance score by destroying targets on time or correctly performing BDA. The base number of points earned for achieving either of these objectives corresponded to the allotted time it took to perform them. Table C-1 shows the number of base points awarded for objective achievement.

Event	Base Points
Target Correctly Destroyed	30
BDA Correctly Performed	45

Table C-1: Base number of points for performing mission objectives [45].

Target priority and difficulty further modified the number of points earned for destroying a target [45]. Target priority referred to the low, medium, and high modifiers assigned to each target within the mission. Target difficulty referred to how difficult it was to destroy a target due to mission re-planning, the number of emergent events near or during the target's TOT, and where in the timeline (start, middle, or end) the target's TOT was scheduled. Each target was assigned a difficulty of low, medium, or hard based upon these factors and previous experimental data on how often operators missed that target. These modifications to target values were made because in real missions, targets would be classified with different priorities and destroying higher priority targets could have tangible effects on an overall mission. Additionally, TOTs occurring during times of high workload or uncertainty would be more difficult to strike; thus, if these difficult targets were destroyed, it would indicate high performance by an operator in managing the mission and achieving its objectives. Table C-2 shows the modified target scores for different combinations of target priorities and difficulties.

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

Table C-2: Modified target scores for different combinations of target priorities and difficulties [45].

Scores for correctly performing BDA were modified only by difficulty, with two classifications (difficult or easy) based upon whether or not the BDA event was affected by mission re-planning [45]. BDA that was scheduled by the subject as part of mission re-planning was considered difficult because the subject had to perceive and act upon a message to include BDA on a target. BDA that was initially scheduled before the mission scenario began was dictated as easy because it was automatic unless altered by the subject. Table C-3 shows modified BDA scores for correctly performed BDA.

BDA Difficulty	Modified BDA Score
Difficult	45
Easy	22.5

Table C-3: Modified BDA scores based upon difficulty [45].

C.2 Penalized Points

Penalty points were deducted from a subject's performance score when he or she performed actions that ran counter to mission objectives. The five types of possible events that incurred penalty points were 1) destroying a target when previously commanded not to, 2) incorrectly performing BDA, 3) having a UAV fired upon in a threat area, 4) arriving at base past the mission time limit, or 5) requesting TOT delays. The fifth event was an addition to the performance score developed in [45]. Table C-4 shows the values of penalty points associated with actions that deter mission objectives.

Event	Penalty Points
Target Incorrectly Destroyed	45
BDA Incorrectly Performed	0
Hit in Threat Area	10
Late Arrival to Base	1 per second per UAV
TOT Delay Request	5

Table C-4: Penalty points associated with actions contrary to mission objectives.

For incorrectly destroyed targets, the penalty deduction value was chosen to be the average score of a correctly destroyed target, 45 points. For incorrectly performed BDA, no penalty was incurred because it was assumed that the wasted time spent performing unnecessary BDA would cause time penalties in future events [45]. The penalty for a UAV hit in a threat area was chosen to be 10 points, as UAVs traversing threat areas were fired upon every ten seconds during the simulation [45]. Because the highest mission objective (Appendix B) was for all UAVs to return to base within the time limit for the mission, the late arrival of a UAV to base past the mission time limit was penalized at 1 point per second per each late UAV [45].

Although requesting TOT delays was an available schedule management capability, past research [45] showed that operators sometimes made excessive requests and thus abused the TOT delay request function. This behavior would have tangible consequences in actual military operations, where excessive requests could cause negative repercussions, such as decreases in an individual operator's performance, saturation of communication lines, and wasting of resources. Thus, a TOT delay request penalized the performance score by one point for each second the request took. Because it took five seconds to receive a response from a TOT delay request, five points were deducted from a subject's performance score for each request, regardless of whether or not it was granted.

Appendix D: 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 E: MAUVE Instructions Experimental Handout Example– Local StarVis Version

Multi-Unmanned Aerial Vehicle Experiment (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 manage a mission timeline while controlling multiple unmanned aerial vehicles (UAVs), given different timeline management decision support visualizations. 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.

Table 1: Color Coding of UAV Actions in MAUVE

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

Objectives

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 = UAV# the waypoint is associated with and Y = waypoint letter
    - Examples: WP-1A, WP-2C
- *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.

Decision Support Visualization

The purpose of the decision support visualization is to assist the user in managing the mission timeline and mitigating two types of schedule problems, a potential high workload area and a late target arrival.

Potential High Workload Area

- A potential high workload area is defined as 2 or more UAVs requiring high workload actions, defined as arming, firing, or battle damage assessment, overlapping with one another on the timeline
- Potential high workload areas are identified on the timeline display by dotted lines outlining the targets contributing to the situation.

Late Target Arrival

- A late target arrival is defined as when a UAV will arrive to a target after its pre-scheduled time on target.

- Late arrivals are identified on the timeline display by the target's black arrival box being located after a target's TOT.

Configural Display
Each UAV has a decision support configural display, shown in Figure 1. A configural display uses emergent features to show timeline issues that currently exist or may exist if the operator requests a schedule change.

- No schedule issues are indicated by only the rectangle (no triangles) being displayed
- Late arrivals and TOT conflicts are represented by the triangles on the left and right halves of the rectangle, respectively.
- Triangles representing high priority targets are on top of the rectangle, medium priority along the side, and low priority on the bottom.
- Gray triangles indicate current timeline issues. Yellow triangles show the "what if" condition and appear after the user selects a target that they might request a TOT delay for from a given list. Split color triangles indicate that the same number of targets of that priority and in that category of timeline issue exists now and in the "what if" condition.
- The height of the triangles indicate how many targets in that priority have that type of schedule issue.

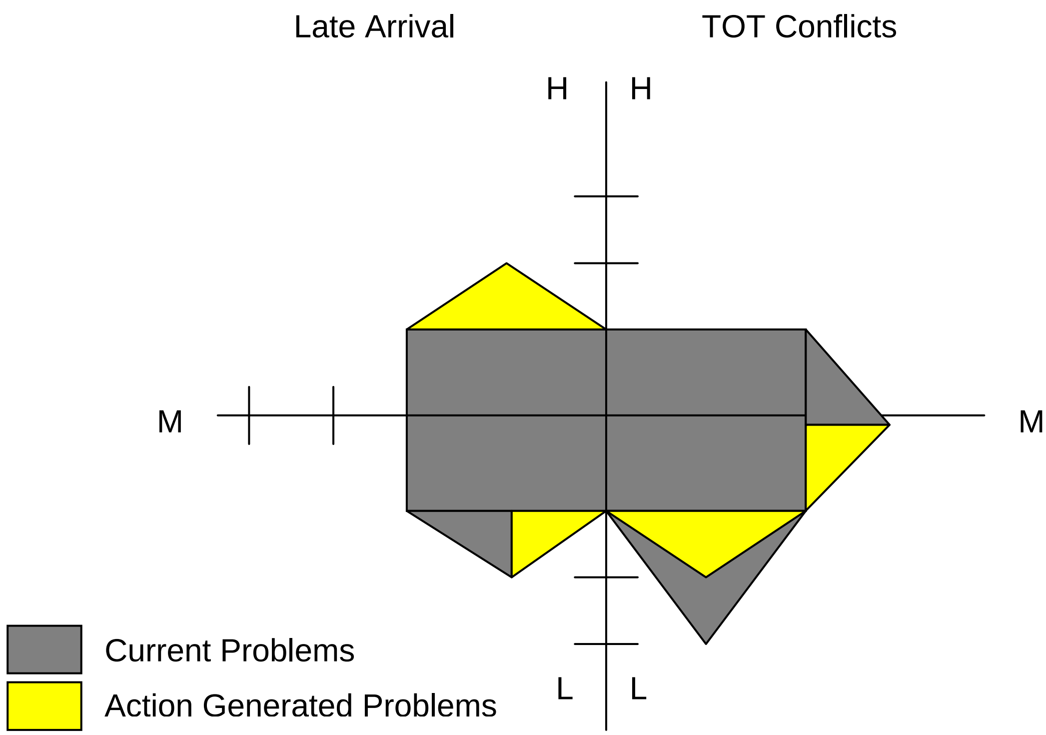


Figure 1: Configural Display Used in Decision Support Visualizations

Using the Visualization
Each UAV possesses a configural display to help the user understand the potential effects of decisions. Each UAV that has timeline issues has a list of targets that either are in a high workload area or will be arrived to late for that UAV only. Before the user checks any of the target selection boxes to show the "what if" condition, each visualization shows the current schedule issues for that UAV. The user can check ONE target selection box per UAV to show what will happen to that SPECIFC UAV's schedule if the selected target's TOT is delayed. More than one UAV can have a target selection box checked. Targets can be deselected to show the current schedule issues condition. Users may use the visualization to "test out" decision alternatives in their goal of managing the mission schedule. Figure 2 shows an example of what the decision support visualizations look like for the UAVs.

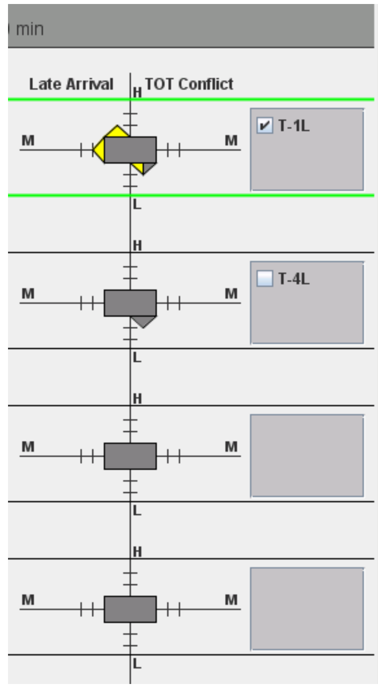


Figure 2: Decision support visualizations for UAVs

Appendix F: Post-Test Questionnaire

MAUVE Post-Test Feedback Form

1. Which features of the MAUVE simulation were the most confusing? Why?

____

____

____

____

2. How did the decision support visualizations help or hinder you in managing your schedule? What aspects of the visualizations were useful? Not useful?

____

____

____

____

3. What improvements to the MAUVE simulation could increase its ease of use?

____

____

____

____

4. What improvements to the MAUVE simulation interface and scenarios could make it more representative of an actual multiple UAV command and control supervision task?

(a) Interface

____

____

____

____

(b) Scenarios

____

____

____

____

5. Please feel free to use the space below to express any comments you have about the simulation you just completed or any other feature of the study (use the back if necessary).

____

____

____

____

____

References

[1] DoD, "DOD Dictionary of Military and Associated Terms," Technical Report Joint Publication 1-02, 12 July 2007.

[2] C. E. Nehme, J. W. Crandall, and M. L. Cummings, "An Operator Function Taxonomy for Unmanned Aerial Vehicle Missions," presented at 12th International Command and Control Research and Technology Symposium, Newport, RI, 2007.

[3] USAF, "The U.S. Air Force Remotely Piloted Aircraft and Unmanned Aerial Vehicle Strategic Vision," *U.S. Air Force Link*, 2005. [Online document]. Accessed 2007 Sept 24. Available: http://www.af.mil/shared/media/document/AFD-060322-009.pdf.

[4] DoD, "Unmanned Aircraft Systems Roadmap 2005-2030," Office of the Secretary of Defense, Washington D.C. 2005.

[5] D. G. Gibbs, "The Predator in Operation IRAQI FREEDOM - A Pilot's Perspective," presented at AIAA Infotech @ Aerospace. Arlington, VA: AIAA, 2005.

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

[7] R. Parasuraman, T. B. Sheridan, and C. 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*, vol. 30, pp. 286-297, 2000.

[8] M. Chanmugavel, A. Tsourdos, R. Zbikowski, and B. A. White, "3D Dublins Sets Based on Coordinated Path Planning for Swarm of UAVs," presented at AIAA Guidance, Navigation, and Control Conference and Exhibit. Keystone, CO: AIAA, 2006.

[9] U. Zengin and A. Dogan, "Cooperative Target Tracking for Autonomous UAVs in an Adversarial Environment," presented at AIAA Guidance, Navigation, and Control Conference and Exhibit. Keystone, CO: AIAA, 2006.

[10] J. Ousingsawat, "Quasi-Decentralized Task Assignment for Multiple UAV Coordination," presented at AIAA Guidance, Navigation, and Control Conference and Exhibit. Keystone, CO: AIAA, 2006.

[11] M. Alighanbari and J. P. How, "Robust Decentralized Task Assignment for Cooperative UAVs," presented at AIAA Guidance, Navigation, and Control Conference and Exhibit. Keystone, CO: AIAA, 2006.

[12] T. Shima and C. Schumacher, "Assignment of Cooperating UAVs to Simultaneous Tasks using Genetic Algorithms," presented at AIAA Guidance, Navigation, and Control Conference and Exhibit. San Francisco, California: AIAA, 2005.

[13] D. Turra, L. Pollini, and M. Innocenti, "Real-Time Unmanned Vehicles Task Allocation with Moving Targets," presented at AIAA Guidance, Navigation, and Control Conference and Exhibit. Providence, RI: AIAA, 2004.
[14] C. Schumacher, P. Chandler, M. Pachter, and L. Pachter, "UAV Task Assignment with Timing Constraints via Mixed-Integer Linear Programming," presented at AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop, and Exhibit. Chicago, IL: AIAA, 2004.
[15] M. Alighanbari, Y. Kuwata, and J. P. How, "Coordination and Control of Multiple UAVs with Timing Constraints and Loitering," presented at American Control Conference. Denver, CO: IEEE, 2003.
[16] T. Furukawa, H. F. Durrant-Whyte, G. Dissanayake, and S. Sukkarieh, "The Coordination of Multiple UAVs for Engaging Multiple Targets in a Time-Optimal Manner," presented at IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, NV: IEEE, 2003.
[17] C. Schumacher, A. DeGlopper, A. Fulford, and D. Kingston, "Cooperative UAV Task Planning for Prosecution of Moving Ground Targets," presented at AIAA Infotech@Aerospace. Arlington, VA: AIAA, 2005.
[18] A. Pongpunwattana and R. Rysdyk, "Real-Time Planning for Multiple Autonomous Vehicles in Dynamic Uncertain Environments," *Journal of Aerospace Computing, Information, and Communication*, vol. 1, pp. 580-604, 2004.
[19] B. R. R. Vandermeersch, Q. P. Chu, and J. A. Mulder, "Design and Implementation of a Mission Planner for Multiple UCAVs in a SEAD Mission," presented at AIAA Guidance, Navigation, and Control Conference and Exhibit. San Francisco, CA: AIAA, 2005.
[20] M. Endsley, "Design and Evaluation for Situation Awareness Enhancement," presented at Human Factors Society 32nd Annual Meeting, Santa Monica, CA, 1988.
[21] A. T. Lefebvre, J. T. Nelson, and T. S. Andre, "Managing Multiple Uninhabited Aerial Vehicles: Changes in Number of Vehicles and Type of Target Symbology," presented at Interservice/Industry Training, Simulation, and Education Conference. Orlando, FL, 2004.
[22] H. A. Ruff, S. Narayanan, and M. H. Draper, "Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles," *Presence*, vol. 11, pp. 335-351, 2002.
[23] J. Nelson, G. Calhoun, and M. Draper, "A Dynamic Mission Replanning Testbed for Supervisory Control of Multiple Unmanned Aerial Vehicles," Air Force Research Laboratory Human Effectiveness Directorate, Wright-Patterson, AFB, OH March 2006.
[24] J. Malasky, L. M. Forest, A. C. Khan, and J. R. Key, "Experimental Evaluation of Human-Machine Collaborative Algorithms in Planning for Multiple UAVs,"

presented at IEEE International Conference on Systems, Man and Cybernetics Hawaii, 2005.
[25] H. A. Ruff, G. L. Calhoun, M. H. Draper, J. V. Fontejon, and B. J. Guilfoos, "Exploring Automation Issues in Supervisory Control of Multiple UAVs," presented at 2nd Human Performance, Situation Awareness, and Automation Technology Conference, Daytona Beach, FL, 2004.
[26] S. L. Howitt and D. Richards, "The Human Machine Interface for Airborne Control of UAVs," presented at 2nd AIAA "Unmanned Unlimited" Systems, Technologies, and Operations - Aerospace, Land and Sea Conference & Workshop, San Diego, CA, 2003.
[27] A. D. White, "Flexible Command and Control of Semi-Autonomous Combat UAVs," presented at AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop, and Exhibit, Chicago, IL, 2004.
[28] G. Barbato, G. Feitshans, R. Williams, and T. Hughes, "Operator Vehicle Interface Laboratory: Unmanned Combat Air Vehicle Controls & Displays for Suppression of Enemy Air Defenses," presented at 12th International Symposium on Aviation Psychology. Dayton, OH, 2003.
[29] S. M. Galster, B. A. Knott, and R. D. Brown, "Managing Multiple UAVs: Are we Asking the Right Question?," presented at Human Factors and Ergonomics Society 50th Annual Meeting, San Francisco, CA, 2006.
[30] S. L. Howitt, D. Richards, and A. D. White, "Supervisory Control of Autonomous UAVs in Networked Environments," presented at AIAA Infotech @ Aerospace. Arlington, VA: AIAA, 2005.
[31] M. V. D'Ortenzio, F. Y. Enomoto, and S. L. Johan, "A Collaborative Decision Environment for UAV Operations," presented at AIAA Infotech @ Aerospace. Arlington, VA: AIAA, 2005.
[32] M. L. Hanson, E. Roth, C. M. Hopkins, and V. Mancuso, "Designing Human System Interfaces for Supervising Multiple UAV Teams," presented at 42nd AIAA Aerospace Sciences Meeting and Exhibit. Reno, NV: AIAA, 2004.
[33] M. L. Cummings and S. Guerlain, "Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles," *Human Factors*, vol. 49, pp. 1-15, 2007.
[34] M. L. Cummings and S. Guerlain, "An Interactive Decision Support Tool for Real-Time In-Flight Replanning of Autonomous Vehicles," presented at AIAA 3rd "Unmanned Unlimited" Technical Conference, Workshop and Exhibit, Chicago, IL, 2004.
[35] M. L. Cummings and P. J. Mitchell, "Automated Scheduling Decision Support for Supervisory Control of Multiple UAVs," *AIAA Journal of Aerospace Computing, Information, and Communication*, vol. 3, pp. 294-308, 2006.

[36] M. L. Hanson, E. Roth, C. M. Hopkins, and V. Mancuso, "Developing Mixed-Initiative Interaction with Intelligent Systems: Lessons Learned from Supervising Multiple UAVs," presented at AIAA 1st Intelligent Systems Technical Conference. Chicago, IL, 2004.

[37] G. Calhoun, H. Ruff, J. Nelson, and M. Draper, "Survey of Decision Support Control/Display Concepts: Classification, Lessons Learned, and Application to Unmanned Aerial Vehicle Supervisory Control," presented at 1st International Conference on Augmented Cognition, Las Vegas, NV, 2005.

[38] B. Shneiderman, *Designing the User-Interface: Strategies for Effective Human-Computer Interaction*, 4th ed. College Park, MA: Pearson Addison Wesley, 2005.

[39] K. B. Bennett and B. Walters, "Configural Display Design Techniques Considered at Multiple Levels of Evaluation," *Human Factors*, vol. 43, pp. 415-434, 2001.

[40] K. B. Bennett, "Graphical Displays: Implications for Divided Attention, Focused Attention, and Problem Solving," *Human Factors*, vol. 34, pp. 513-533, 1992.

[41] C. D. Wickens and C. M. Carswell, "The Proximity Compatibility Principle: Its Psychological Foundation and Relevance to Display Design," *Human Factors*, vol. 37, pp. 473-494, 1995.

[42] J. J. Gibson, *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.

[43] M. A. Buttgieg and P. M. Sanderson, "Emergent Features in Visual Display Design for Two Types of Failure Detection Tasks," *Human Factors*, vol. 33, pp. 631-651, 1991.

[44] P. Sanderson, "The Human Planning and Scheduling Role in Advanced Manufacturing Systems: An Emerging Human Factors Domain," *Human Factors*, vol. 31, pp. 635-666, 1989.

[45] P. J. Mitchell, "Mitigation of Human Supervisory Control Wait Times through Automation Strategies," M.S. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2005.

[46] M. L. Cummings and S. Guerlain, "Using a Chat Interface as an Embedded Secondary Tasking Tool," presented at Human Performance, Situation Awareness and Automation Technology II Conference, Daytona Beach, FL, 2004.

[47] S. Hart and L. Staveland, "Development of the NASA-TLX: Results of Empirical and Theoretical Research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds., 1.0 ed. Amsterdam: North Holland, 1988, pp. 139-183.

[48] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, vol. 37, pp. 32-64, 1995.

[49] A. Seybold, "Design and Analysis of a Reduced Complexity Decision Support Visualization for Multiple UAV Scheduling," S.B. thesis, Cambridge, MA, 2007.

[50] A. S. Brzezinski, A. L. Seybold, and M. L. Cummings, "Decision Support Visualizations for Schedule Management of Multiple Unmanned Aerial Vehicles," in *AIAA Infotech @ Aerospace*. Rohnert Park, CA: AIAA, 2007.