

Operator Scheduling Strategies in Supervisory Control of Multiple UAVs

M.L. Cummings*
Massachusetts Institute of Technology
Humans and Automation Lab

P.J. Mitchell
Power Generation Technology
General Electric Energy

Abstract

The application of network centric operations to time-constrained command and control environments will mean that human operators will be increasingly responsible for multiple simultaneous supervisory control tasks. One such futuristic application will be the control of multiple unmanned aerial vehicles (UAVs) by a single operator. To achieve such performance in complex, time critical, and high risk settings, automated systems will be required both to guarantee rapid system response as well as manageable workload for operators. Through the development of a simulation test bed for human supervisory control of multiple independent UAVs by a single operator, this paper presents recent efforts to investigate workload mitigation strategies as a function of increasing automation. A human-in-the-loop experiment revealed that under low workload conditions, operators' cognitive strategies were relatively robust across increasing levels of automated decision support. However, when provided with explicit automated recommendations and with the ability to negotiate with external agencies for delays in arrival times for targets, operators inappropriately fixated on the need to globally optimize their schedules. In addition, without explicit visual representation of uncertainty, operators tended to treat all probabilities uniformly. This study also revealed that operators that reached cognitive saturation adapted two very distinct management strategies, which led to varying degrees of success. Lastly, operators with management-by-exception decision support exhibited evidence of automation bias.

Keywords: multiple unmanned aerial vehicles, human supervisory control, levels of automation

* Corresponding author, 77Massachusetts Ave, 33-305, Cambridge, MA 02139, missyc@mit.edu, 1-617-252-1512, 1-617-253-4196 (fax).

1.0 Introduction

Human supervisory control (HSC) occurs when a human operator monitors a complex system and intermittently executes some level of control on the system through some automated agent. Because HSC is not typically a continuous task, HSC operators can typically time-share cognitive resources with other tasks. This task time-sharing is critical for the success of network-centric operations (NCO) which require operators to process information from a variety of sources, often within a short period of time. One such domain where HSC and NCO will intersect in the future will be in unmanned vehicle operations. While many operators are presently needed to control a single unmanned aerial vehicle (UAV), as technology and autonomous control improve, automation will handle lower level tasks, thus enabling the supervisory controller more time to control a greater number of UAVs. The key to achieving this one-controlling-many goal will be the development of automated control architectures that complement human decision-making processes in time-critical environments.

A significant cognitive control issue for humans managing multiple vehicles is that they must synthesize voluminous data from a network of sensors and vehicles, and then execute decisions in real-time, often with high-risk consequences under significant uncertainty. In time-pressured scenarios like those expected in NCO, efficiently allocating attention between a set of dynamic tasks becomes critical to both human and system performance. In the future vision of allowing a single operator to control multiple unmanned vehicles (which could be on land, in the air, or under water), it is not well understood how operators will manage multiple vehicles, what kind of decision support can complement operators, and how human cognitive limitations will impact overall system effectiveness. To this end, this paper will discuss recent efforts to investigate operator workload mitigation strategies as a function of increasing automation in supervisory control of multiple independent and homogenous UAVs.

2.0 The Need for Nested Automation in Supervisory Control of Multiple Vehicles

In terms of operator task loading, UAV control is hierarchical, as represented in Figure 1. The inner loop represents a basic guidance and motion control loop for a UAV which is the most critical loop that must obey aerodynamic constraints or the UAV will not remain airborne. The second loop,

the navigation loop, represents the actions that some agent, whether human or computer-driven, must execute to meet mission constraints such as routes to waypoints, time on targets, and avoidance of threat areas and no-fly zones. The outermost loop represents the highest levels of control, that of mission and payload management. For typical UAV missions such as intelligence, surveillance, and reconnaissance (ISR), sensors must be monitored and decisions made based on the incoming information to meet overall mission requirements. Finally, the system health and status monitoring loop represents the continual supervision that must occur, either by a human or automation or both, to ensure all systems are operating within normal limits. As represented by the nested loops, if the inner loops fail, then the higher or outer loops will also fail.

The dependency of higher loop control on the successful control of the lower loops drives human limitations in control of multiple unmanned vehicles. For a human operator to successfully control multiple UAVs, the innermost loop must be fully automated so that operators' cognitive resources can be focused on higher loop control. If humans must continuously interact in the guidance and motion control loop (fly the UAV), the cost is high because this effort requires significant cognitive resources, and what little spare mental capacity is available must be divided between the navigation and payload management control loops. For this study, the guidance and motion control loop is assumed to be fully automated, with an intermediate level of automated navigation support. For the mission and payload management loop, increasing degrees of automation for payload scheduling support were used which represent the primary independent variable of interest.

<<Insert Fig. 1>>

2.1 Levels of Automation

While automating significant aspects of NCO is necessary in order to handle large quantities of information and also so that information sharing can be both quick and comprehensive, what to automate and to what degree to automate a process/system is a central question in the design of NCO systems such as those that will supervise multiple UAVs. Sheridan and Verplank [1] outlined a scale from 1-10 where each level represented the machine performing progressively more tasks than the

previous one, as shown in Table 1. Human interaction with automation represents a range of intermediate levels from 2-6 on this scale. For routine operations, higher levels of automation (LOAs) in general result in lower workload, while the opposite is true for low levels of automation [2]. If the LOA is too high, HSC operators could experience mental skill degradation and loss of situational awareness due to lack of automation transparency, complexity, and inadequate feedback. However, since automation will be critical to the successful control of multiple vehicles, if the LOA is too low, operators could experience working memory overload in routine tasks under time pressure and greater human interdependency and chaos when something fails [3, 4].

A few studies have investigated levels of automation in the context of multiple UAV supervisory performance. Ruff et al. [5, 6] examined the effects of levels of automation and decision-aid fidelity in human interaction with up to 4 UAVs. They found that a medium level of automation called management-by-consent, which corresponds to an automation level of 5 on the scale of Table 1, had significant advantages over manual control (Level 1, Table 1). However, results were mixed for the management-by-exception (Level 6, Table 1) supervisory control schemes. In the first study, the moderate LOA produced the highest levels of operator situation awareness (SA) and performance, however in a subsequent study there was no difference.

One drawback to these two studies was the lack of distinction between the LOAs across the different control loops as depicted in Figure 1. LOAs can vary across the embedded control loops, and a general assignment of a single LOA across the three loops makes it difficult to determine how to effectively model and intervene to free specific cognitive resources, either from an automation or decision support perspective. Wickens et al. [7, 8] demonstrated that automating the guidance and motion control loop reduced operator workload by freeing cognitive resources for other tasks, and that some automation for the navigation and mission management control loops was helpful in reducing operator workload. However, they also did not distinguish between levels of automation for the navigation and mission management loops, thus it is not clear how the levels of automation in each loop affected the other and the overall outcome of the mission. To prevent this confound, in this study we held constant the automation in the inner two loops and only examined how varying the levels of

automation for the mission and payload management loop affect operator performance in a schedule management task.

<<Insert Table 1>>

3.0 MAUVE: The Experimental Test Interface

In order to study how levels of automation affect operators strategies for schedule maintenance and payload management, a dual screen simulation test bed named the Multi-Aerial Unmanned Vehicle Experiment (MAUVE) interface was developed (Figure 2). In this simulation, users took on the role of an operator responsible for supervising four UAVs tasked with destroying a set of time-sensitive targets in a suppression of enemy air defenses (SEAD) mission. The UAVs launched with a pre-determined mission plan that came from a pre-determined air tasking order (ATO), so initial target assignments and routes were already completed. The operator's job was to monitor the progress of the UAVs, 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.

The UAVs supervised by participants in this MAUVE experiment were capable of six high-level actions in the simulation: traveling enroute to targets, loitering at specific locations, arming payloads, firing payloads, performing battle damage assessment, and returning to base, generally in this order. Battle damage assessment (BDA, 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. BDA in MAUVE is semi-automated in the sense that operators are responsible for scheduling BDA in advance, but the UAV performs it automatically after firing, if scheduled.

<<Insert Fig. 2>>

The left-hand side of the MAUVE interface in Figure 2 is known as the navigation display, and it consists of a mission time window, map display, and a mission planning and execution bar. The map display contains iconic UAVs, targets, waypoints, and loiter patterns. The map display also represents

threat areas which can dynamically change over time. Routing changes such as the addition or removal of waypoints, loiter points, or targets can be accomplished using the mission planning and execution bar to the left of the map, but are only required for real-time re-planning as a result of unexpected events. The most critical control function accessible to the operator in MAUVE is the “Request TOT Delay” button which allows operators limited opportunities to manipulate the time-on-targets (TOTs) for those targets assigned to a particular UAV. TOTs are critical events in that they are scheduled in advance as part of a coordinated strike with multiple agencies, and often they are highly constrained due to operational needs. For example, a UAV may be ordered to drop a weapon in a very narrow window of time because friendly forces may be in the designated area at all other times. In MAUVE, operators can request a TOT delay for a given target for two reasons: 1) According to the current mission plan, they are predicted to arrive late to a particular target and therefore will miss their deadline, or 2) for workload purposes, i.e., if an operator feels they need to spread out a future potentially high workload. The probability of approval for a TOT change in MAUVE is a function of how far in advance of the deadline the request is made, as would likely be the case in true military situations. Thus, if a TOT deadline is immediately approaching, the chance of approval is zero, but nearly 1.0 when requested 15 minutes or more in advance.

The right-hand side of the MAUVE simulation in Figure 2 provides decision support, and consists of a UAV status window, chat box, UAV health and status updates, and the decision support window. The bottom left of this display has a text-based communication tool known as a chat box that contains a time history of all human communication interactions. The chat box is included because it is an established method of communications in current day military command and control scenarios [9], and is an externally valid embedded secondary workload and situation awareness measurement tool [10]. One message that is particularly important to operators is notification that a TOT request is accepted or denied. The bottom right of the decision support display contains a UAV health and status notification window which separates human communications in the simulation from system communications, and only contains messages from individual UAVs. More interface specifics are discussed at length elsewhere [11].

The decision support in the top right of the decision support display was the primary independent variable for the experiment discussed in this paper. The decision support tool was created to simplify standard air tasking order (ATO) data and combine this into a single interface with up-to-date mission planning information. An ATO generally provides a schedule of events for a strike mission and those required resources needed over a period of hours and/or days to support this mission. Because of the complexity and dynamic nature of an ATO (see Figure 3 for an example of an actual ATO, courtesy of the U.S. Navy), some level of automated decision support is likely to aid in mitigating operator workload, however, it is not clear what level of automation provides the most improvement in schedule maintenance and reduction of operator workload while avoiding negative side-effects, such as a loss of situation awareness. Therefore, four versions of the decision support were created at different levels of automation and structured so that higher levels of decision support expanded upon the features found in lower levels while still retaining all of the functionality and basic information content from previous levels. Thus there were four possible forms of decision support in this experiment that roughly correspond to levels 1, 2, 4, and 6 (Table 1), termed manual, passive, active, and super-active respectively.

The manual LOA level of decision support (Figure 4a) presents all required ATO and mission planning information in a text-based table format. Even though this was considered the least helpful of all the decision support tools, it represents an improvement over what operators must use in current operations (Figure 3). Current TOT windows and estimated times of arrival (ETA) for up to the next four targets in the scheduling queue are presented for easy comparison. ETAs for arrival at home base and the next waypoint or navigation point on the current route segment (if applicable) are also given. Further assistance is provided to the user through the “Next Expected Action” column, which tells the user what they should be doing next and at what time, according to the preplanned ATO. However, this information is updated dynamically to reflect changing ATO requirements and mission planning changes initiated by the user.

<<Insert Fig. 3>>

The passive LOA (Figure 4b) assimilates all of the ATO and current mission information contained in the manual level and transforms it into a horizontal timeline format color coded by action. The major difference between the passive and the manual level is schedule integration through a representation aiding graphical format which allows users to perceptually assimilate multivariate pieces of information quickly instead of requiring them to use working memory to piece together all the relevant information. In the graphical timeline, target ETAs are represented by black rectangles, and waypoint, loiter point and base arrival times are marked by black triangles on the bottom of each timeline. Static ATO elements such as target TOT windows, arming windows, and BDA are represented by red, yellow and brown blocks of time at the appropriate times. The blue segments of the timeline represent the enroute phases of flight while the green indicates that the UAVs are headed back to the home base.

The active LOA for this study (Figure 4c) uses the same horizontal timeline format as the passive automation level, but provides additional assistance from the computer in that an algorithm searches for periods of time in the schedule that it predicts will cause high workload for the operator. The computer identifies a high workload area, or “bottleneck” as a period of time during which multiple UAVs are scheduled to be simultaneously executing mission critical actions, defined as arming, firing, or performing BDA, which the operator would have difficulty executing concurrently. The automation draws attention to these areas of concern by a reverse shading technique, in which the “bottlenecks” are highlighted while the rest of the timeline’s colors are muted, but still visible. This design technique was included to allow operators to maintain situation awareness of current events while understanding what the automated predictions were for future events.

In addition to identifying areas of high workload through the reverse shading technique, the computer also recommends a course of action to alleviate the high workload areas, such as moving a particular TOT. In this version of MAUVE, the automation only makes locally optimal recommendations, and the algorithm is not globally optimal. For example, in the case of a TOT conflict between two UAVs, the automation highlights the conflict between the two UAVs, but makes no recommendation as to which UAV should be moved in terms of higher order mission objectives. Thus if an operator follows the computer’s recommendation to relieve a high workload area for a

single UAV, this will remove that particular schedule conflict, but could potentially create another in the downstream process.

The super-active LOA (Figure 3d) also builds upon the passive level graphical timeline, but instead of making recommendations to the operator as in the active LOA, a management-by-exception approach is taken whereby the computer automatically executes the arming and firing actions for all UAVs at each target, when the pre-established rules of engagement for such actions are met. For example, in order to fire, a UAV must be at the pre-designated target, already armed and within the preset TOT. In this highest level of automation, the automation executes all tasks, but the operator is still responsible for determining if the arming and firing actions are appropriate, as well as re-planning actions and manipulating routes to ensure the UAVs arrive at the correct targets on time. In the super-active condition, up to 30 seconds in advance before every arming and firing action, exception boxes appear to the right of each UAV timeline that allow the operator to potentially veto these actions. These boxes are color-coded to indicate which action the UAV is preparing to perform: red for firing and yellow for arming, which matches the related phases on the graphical timeline. This level of decision support is termed super-active because the automation performs all of the mission critical execution actions for the user.

<<Insert Fig. 4>>

4.0 Method

In order to investigate human performance and cognitive strategies in the management of dynamic scheduling tasks for supervisory control of multiple UAVs, an experiment with the MAUVE simulation interface was conducted. The goal of the experiment was to determine how increasing levels of automation and changes in workload would affect performance and operator situational awareness, as well as give insight to the strategies that operators used to manage multiple vehicles in dynamic situations.

4.1 Apparatus, Participants, and Procedure

Training and testing was conducted on a four screen system called the multi-modal workstation (MMWS), envisioned by the U.S. Navy to be the futuristic workstation of operators engaged in the management of multiple tasks towards a common goal. This experimental set-up was in keeping with the typical MMWS configuration as described by Osga et al. [12]. Two screens displayed the dynamic information as depicted in Figure 2, and the additional two screens contained prioritized objectives for the scenarios and color coding for UAV actions. During testing, all mouse clicks and message box histories, including incoming and outgoing messages, were recorded by software. In addition, screenshots of both simulation screens were taken approximately every two minutes. All four UAV locations were recorded every 10 seconds and a whenever a UAV's status changed, the time and change made were noted in the data file.

A total of 12 participants took part in this experiment, 10 men and 2 women. Participants were recruited based on whether they had UAV, military and/or pilot experience. The subject population consisted of a combination of students, both undergraduates and graduates, as well as local reserve officer training corps (ROTC) and active duty military personnel. All were paid \$10 an hour for their participation. In addition, a \$50 incentive prize was offered for the best performer in the experiment. The age range of participants was 20 – 42 years with an average age of 26.3 years.

Participants were briefed that they had two main objectives in this experiment: 1) To guide each UAV so that together, all UAVs under their supervision properly executed the required goals of the current ATO, which could change over time, and 2) To answer periodic questions about the situation from commanders. All participants received between 90 and 120 minutes of training until they achieved a basic level of proficiency in monitoring the UAVs, redirecting them as necessary, executing commands such as firing and arming of payload at appropriate times, and responding to online instant messages. Following training, participants tested on two consecutive 30 minute sessions which were randomized and counter-balanced to prevent a possible learning effect. Each simulation was run approximately four times faster than real time so an entire strike could take place over 30 minutes, instead of several hours as is commonplace in actual strikes.

4.2 Experimental Design

Two independent variables were of interest in this experiment: 1) level of automated decision support (Figure 4) and 2) level of re-planning. The re-planning events consisted of unplanned contingencies such as emergent threat areas and targets, and new tasking from superiors that added or deleted BDA tasks from the overall mission, as well as system failures that might have required a UAV to return to base unexpectedly. For the re-planning experimental factor, low and high levels of schedule re-planning were investigated in which 7 or 13 unexpected events occurred, respectively. Under high re-planning, groups of 2 or 3 re-planning events occurring within 60 seconds of each other. The level of automation was a between-participants variable and the level of re-planning was a within-participants variable, so participants were randomly assigned to a LOA factor level but experienced both re-planning conditions. The dependent variable was a performance score, which measured how well participants met the numerous objectives for a test session. The performance score was a product of the targets correctly destroyed, including their priority and difficulty level, and number of times BDA was correctly performed. Operators were penalized for erroneously firing on incorrect targets and penalties were also assessed for hits taken by UAVs in threat areas, as well as late arrival of UAVs at base. In addition, situation awareness was measured both objectively and subjectively which will be discussed more in depth below.

5.0 Results

A 2x4(3) repeated measures linear mixed model was used to analyze the performance scores. Three participants were nested within each of the four automation levels, and every subject saw both low and high levels of re-planning. Age was used as a covariate in the analysis, and for all reported results $\alpha = 0.05$. Figure 5 shows the average performance scores for each experimental condition within one standard error. Level of re-planning was significant ($F(1, 9.9) = 19.40, p = 0.001$) while level of automation was marginally significant ($F(3, 10.8) = 3.04, p = 0.076$). There was no significant interaction between the factors. Pair-wise comparisons yielded no difference in the levels of automation across the low re-planning level. For the high re-planning level, manual, passive, and

super-active results were statistically no different, however, active automation was statistically lower than the others.

Based on the results from the performance score analysis, more investigation was needed to determine why the active level of automation produced such poor scores, especially in comparison with the lower levels of automation. The significant driver of this problem was found to be operators' use of the "TOT Delay" function as described in section 3.0. Through requesting TOT delays, operators could manipulate target deadlines to spread out their workload and/or make deadlines they otherwise would have missed due to inadequate prior planning and/or execution. The underlined number associated with each mean in Figure 5 is the average count of TOT delay requests for each level of re-planning and automation. The count of TOT delays requested was examined using a Kruskal-Wallis non-parametric analysis which showed that the number of TOT delays requested was marginally significant between automation levels ($p = 0.096$). For both high and low workload re-planning levels, the difference in number of TOT delay requests was driven by a lower number for the manual level of automation (Wilcoxon Rank Sum Test, $p = 0.015$), and when compared just across the high level of re-planning, there were a significantly higher number of TOT requests in the active level ($p = 0.065$).

<<Insert Fig. 5>>

As previously mentioned, situation awareness was measured by assessing the accuracy of embedded SA questions introduced through the instant messaging "chat box." By definition, situation awareness has three elements which include 1) the perception and 2) the comprehension of current events, and lastly 3) the projection of the likely outcomes from elements 1 and 2 [13]. A Wilcoxon Signed Rank Test for level 2 SA scores showed a marginally significant difference between levels of re-planning ($p = 0.065$) and a Kruskal-Wallis test showed a marginally significant difference between automation levels ($p = 0.084$), with the highest level 2 SA occurring in the super-active condition. Participants in the super-active level had fewer actions to perform, and as a consequence had more

time to observe events on the screens, thus they had more time to perceive and comprehend current events.

However, operators using the super-active automation gained no apparent benefit for level 3 SA (future projection) with the extra time as measured by the embedded SA questions introduced through the chat box. A Wilcoxon Signed Rank test for level 3 SA scores demonstrated lower SA under the high re-planning scenario ($p = 0.004$), but there was no effect of level of automation. However, the use of questions, either in real-time or post-hoc, to measure SA may not always be a valid measure of SA [14], and often an overall measure of performance can demonstrate an aggregate level of SA [15]. Thus a performance-based error of commission SA metric was also included in this experiment which was defined as the number of weapons incorrectly released on the wrong targets. This performance-based metric represents how well operators understand both current and future events, thus is an aggregate measure of SA, but it also has a direct parallel in real military conflicts. Occasionally targets would be removed from the ATO which was announced through the chat box, and there were nominally three opportunities for errors of commission in the high re-planning scenario and two in the low re-planning scenario. However, the number of potential opportunities for error varied by participant, as it was contingent on the operator guiding the appropriate UAVs to the targets of interest in the first place. A Wilcoxon Signed Rank test showed there to be a significantly higher number of targets erroneously destroyed under the high re-planning condition as compared to the low ($p = 0.034$). In fact, while seven targets were erroneously destroyed under high workload, only 1 target was under low workload. A non-parametric Kruskal-Wallis test showed no significant difference in errors of commission between levels of automation. However, while not statistically significant, participants in the manual and super-active levels of automation committed more errors of commission than those in the passive and active levels (Figure 6).

<<Insert Fig. 6>>

6.0 Discussion

The re-planning independent variable was included to determine the effect of exogenous task complexity on operator performance and was significant across both performance and situation awareness scores. Under low external workload, regardless of the level of automated decision support, all operators were able to effectively manage the dynamic scheduling problem, and had higher situation awareness. However, under high re-planning, performance generally dropped regardless of the level of automation. However, unfortunately for all levels of automation, and in particular for active automation, operators adapted to the high workload by requesting more TOT delays. This exacerbated already high levels of workload because operators had a more difficult time assessing when the appropriate time would be to request delays, which resulted in a condition we term “cognitive saturation.” This represents the inability for operators to correctly assimilate data from multiple sources, incorrectly weigh uncertainty in the solution space, and fail to correctly prioritize current tasks.

6.1 Cognitive Saturation

During testing, a measurable point of cognitive saturation was observed when participants were overwhelmed, which was generally caused either by exogenous factors (the frequent re-planning events), or endogenous factors (a loss of SA). This saturation point was only seen in the high re-planning test sessions, as operators in the low re-planning scenario had enough rest periods to allow for recovery without serious loss of system performance. However, once overloaded, operators’ behavior followed two distinct patterns: 1) they began to supervise fewer vehicles, ignoring one or more entirely, or two) they attempted to supervise all vehicles, missing a majority of their deadlines while they tried to do so. Figure 7, which plots actual UAV paths against the optimal paths over the course of two high re-planning trials, illustrates both strategies.

In Figure 7a, the operator employed a cognitive shedding strategy whereby his interaction effort was focused on a smaller subset of vehicles than originally tasked. This operator chose to focus on only two vehicles at a time and while three targets were missed and one UAV experienced a slight incursion into a hazardous area, the operator’s strategy to focus on a smaller number of vehicles led to superior performance as compared to the operator’s illustrated in Figure 7b. Instead of using a

cognitive shedding strategy in terms of focusing on a smaller subset of vehicles, in Figure 7b the operator attempted to control all four simultaneously. This resulted in leaving one UAV loitering at a target for nearly five minutes before he realized it had stopped executing the mission plan. The operator, clearly saturated with this effort, missed six targets and rerouted a UAV through a threat area. The operator in Figure 7b eventually adapted a cognitive shedding strategy and elected to send two UAVs directly home instead of managing them but by this time, the mission was severely compromised.

<<Insert Fig. 7>>

6.2 The Impact of the Increasing Levels of Automation

This study demonstrated under low workload levels, operator performance was fairly robust across all levels of automation. The different coping strategies discussed in the previous section were not exhibited until workload was increased, causing operators to interact more frequently with all UAVs. In order to prevent some of the poor coping strategies exhibited in Figure 7 as a function of high workload, some level of automation will be needed to alleviate the workload, but to what degree remains unclear. While Ruff et al. (2002) found that a management-by-consent LOA (corresponding most closely to the active level in this study), had performance advantages over manual and management-by-exception (equal to our super-active) conditions for multiple UAV supervision, our results demonstrated that management-by-consent in terms of this specific active decision support produced the worst performance.

This study demonstrates that even a slight change in the information presentation and intelligent autonomy of a decision aid can produce significant and unexpected results. Specifically the active automation level in this study, designed to aid operators in schedule management, caused significant operator performance degradation. Primarily, the performance decrement under the active level of automation can be attributed to participants' inappropriate use of the "Request TOT Delay" function which added to their workload, especially under the high re-planning condition. The salience of the high workload prediction is likely an additional and related reason for this degraded performance. The

reverse shading technique was used to suggest where potential workload bottlenecks might occur in the future, with those predictions in the next 5 minutes highly likely but those 10-15 minutes away much less likely. Under low workload, operators were able to correctly assimilate this knowledge but when stress and workload increased, operators treated the likelihood of schedule bottlenecks as having a uniform probability. With the intelligent aiding provided in this experiment, operators were not able to accurately assess future workload states, which also affected their ability to prioritize their current workload, and led to significantly degraded performance. Operators with much less automated decision support and no intelligent aiding exhibited superior performance under high workload conditions, highlighting the fact that even a relatively minor visualization change (the reverse shading) in concert with future workload predictions can produce dramatically unintended results.

6.3 Automation Bias

Despite having the highest level of perception/ comprehension SA as measured by the embedded situation awareness questions, participants in the super-active automation condition erroneously destroyed more targets than those under passive and active automation. This suggests that participants with management-by-exception automation suffered from complacency, as they were content to let the computer take care of the arming and firing actions without close monitoring and cross-checking the release of weapons with the updated ATO. This may be a case of automation bias, which occurs when operators do not seek disconfirming evidence for computer-generated recommendations, and this bias has been demonstrated before in other time-critical command and control applications [16].

If the costs of an erroneous automated action are high, such as what would be expected in erroneous release of a weapon, management-by-exception systems may not be a feasible option, despite the fact that operators can achieve higher levels of overall performance. However, for UAV missions that do not involve high-risk critical events such as those that are more passive in nature such as reconnaissance missions, higher levels of automation may be more appropriate.

7.0 Conclusion

In order to study how levels of automation affect the UAV knowledge-based mission and payload management control loop from a human supervisory control perspective, a study was conducted with a simplified simulation of multiple UAVs operating independently of one another. The goal was to determine how increasing levels of automation affected operator performance in order to identify possible future automation strategies for multiple UAV scheduling. The following cognitive strategies were noted in this experiment:

- 1) Under low workload conditions, operators' cognitive strategies were relatively robust across increasing levels of automated decision support.
- 2) Once overloaded, operators who adapted a temporary local control strategy, i.e., only concentrating on a reduced number of UAVs for short periods of time performed better than those operators who attempted to time-share attention across all UAVs.
- 3) Operators with predictive decision support could adequately weigh uncertainty and fixated on possible future areas of high workload, to the detriment of the current task.
- 4) Operators with management-by-exception decision support exhibited evidence of automation bias.

The success of single operator control of multiple autonomous vehicles and in general, network-centric operations depends on operators' abilities to process information from a variety of sources, often within a short period of time. Thus higher levels of automation will be needed to accomplish this, but as this research highlights, the introduction of intelligent decision aiding can often produce unexpected results. This research highlights the need to understand the impact of intelligent aiding on operator cognitive strategies, particularly in the control of multiple unmanned vehicles which will require significant levels of autonomy. In addition, more research is needed to identify shifts in operator strategies as a function of automated decision support and increased workload, and how best to design a system that can support the operator while also ensuring mission goals are achieved.

Acknowledgments

This research was sponsored by Boeing Phantom Works. We would also like to thank Dr. Tom Sheridan for his assistance and invaluable insight.

References

- [1] T. B. Sheridan and W. L. Verplank, "Human and Computer Control of Undersea Teleoperators," MIT, Cambridge, Man-Machine Systems Laboratory Report 1978.
- [2] D. B. Kaber, M. R. Endsley, and E. Onal, "Design of Automation for Telerobots and the Effect on Performance, Operator Situation Awareness and Subjective Workload," *Human Factors and Ergonomics in Manufacturing*, vol. 10, pp. 409-430, 2000.
- [3] 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.
- [4] C. E. Billings, *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.
- [5] 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.
- [6] 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 Conference (HPSAA II), Daytona Beach, FL, 2004.
- [7] C. D. Wickens, S. Dixon, and D. Chang, "Using Interference Models to Predict Performance in a Multiple-Task UAV Environment - 2 UAVs," Aviation Human Factors Division, Institute of Aviation, University of Illinois at Urbana-Champaign, Savoy, IL AHFD-03-9/MAAD-03-1, April 2003 2003.

- [8] C. D. Wickens and S. Dixon, "Workload demands of remotely piloted vehicle supervision and control: (I) Single vehicle performance (ARL-02-10/MAD-02-1)," University of Illinois, Aviation Research Lab, Savoy, IL 2002.
- [9] M. L. Cummings, "The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations," presented at Adaptive Displays Conference, Los Angeles, CA, 2004.
- [10] M. L. Cummings and S. Guerlain, "Using A Chat Interface As An Embedded Secondary Tasking Tool," presented at 2nd Annual Human Performance, Situation Awareness, and Automation (HPSAA) Conference, Daytona Beach, FL, 2004.
- [11] P. J. Mitchell, M. L. Cummings, and T. B. Sheridan, "Management of Multiple Dynamic Human Supervisory Control Tasks," MIT Humans and Automation Laboratory, Cambridge, MA HAL2005-02, May 2005.
- [12] G. Osga, K. Van Orden, N. Campbell, D. Kellmeyer, and D. Lulue, "Design and Evaluation of Warfighter Task Support Methods in a Multi-Modal Watchstation," SPAWAR Systems Center, San Diego, CA Technical Report 1874, May 2002 2002.
- [13] M. R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, vol. 37, pp. 32-64, 1995.
- [14] M. R. Endsley, "Theoretical Underpinnings of Situation Awareness: A Critical Review," in *Situation Awareness Analysis and Measurement*, M. R. Endsley and D. J. Garland, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- [15] A. R. Pritchett and R. J. Hansman, "Use of Testable Responses for Performance-Based Measurement of Situation Awareness," in *Situation Awareness Analysis and Measurement*, M. R. Endsley and D. J. Garland, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2000, pp. 189-209.
- [16] M. L. Cummings, "Automation Bias in Intelligent Time Critical Decision Support Systems," presented at AIAA Intelligent Systems, Chicago, IL, 2004.

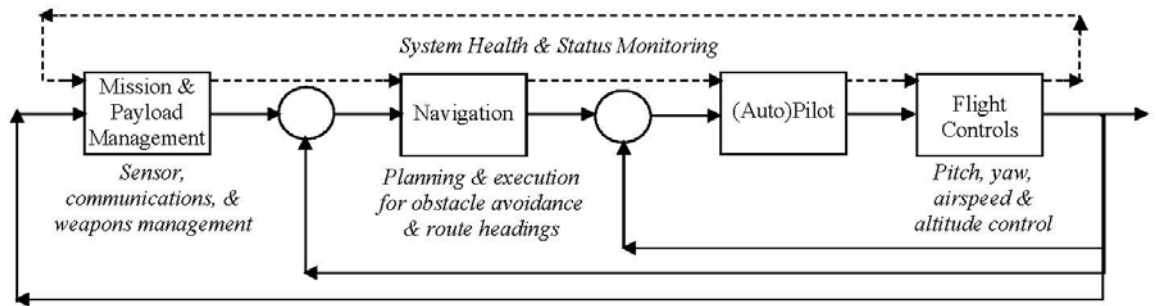


Figure 1: Embedded Control Loops for UAV Operation

Table 1: Levels of Automation [1]

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

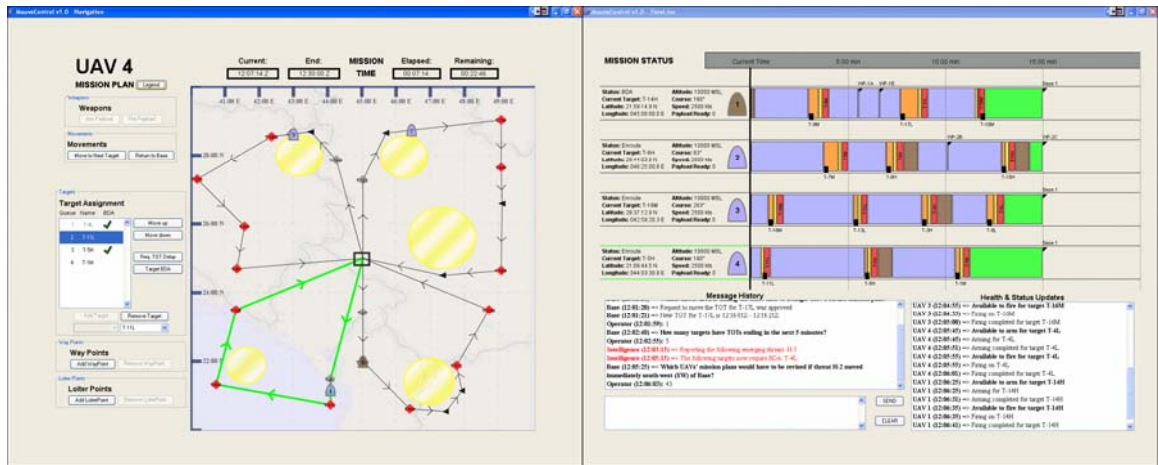


Figure 2: The MAUVE Dual Screen Interface

OPER/DESERT STORM//
 MSSIO/MISREP/CTG 154,11/059/FEB//
 REF/A/RMG/USCENTAF HQ/141356ZFE891//
 AMPN/AIR TASKING ORDER//
 NARR/KTO KILLBOX AG6, CHEMICAL WEAPONS AND TRUCKS//
 HEADING/AIR OPERATIONS//
 NSNID/MSNND'S 0341I, 0343I, 0345D, 0351P, AND 0353Y//
 ENTASK/CV61/152315Z/160158Z/CV61//
 HEADING/TARGET STRUCK-SIGHTED//
 1MSWTGT
 /DE/TGT-ID /LOCATION /DMP1
 /01/ -/294248N0473706E/MILITARY TRUCKS
 /02/ -/294300N0473006E/TRUCKS WITH POSS WEAPONS
 /03/ -/294300N0473005E/PROBABLE TRUCKS
 /04/ -/294236N0473706E/2 MILITARY TRUCKS//
 AMPN/THE STRIKE PACKAGE FLEXED TO THIS TARGET IN KTO KILLBOX AG6
 PRIOR TO LAUNCH. THE TARGET AREA CONSISTED OF PROBABLE CHEMICAL
 WEAPONS AND TRUCKS,//
 1TGTDM

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

MID	INFO	OPS	STROPS	INTEL	CDC	NAV	ROUTE
	CO	XO	ASA	MPC			SLRRAC
	EWU	COMMO					
CCG5	INFO	02-ADCCOP					ROUTE 009//
	01-00	01-01	01-02	01-03	01-04	01-05	01-06
	01-002	01-003	01-004	01-31	01-32	01-33	01-34
	01-35	01-36	01-TOTAL	MID CO	MIDMPC	MIDCOM	

PRH 15295 COPIES 019/034 GIAA 00186 (12) CTG ONE FIVE FOUR PT
 TOR 160430Z TOP 160430Z RUMGSG 0551 160350Z FEB 91

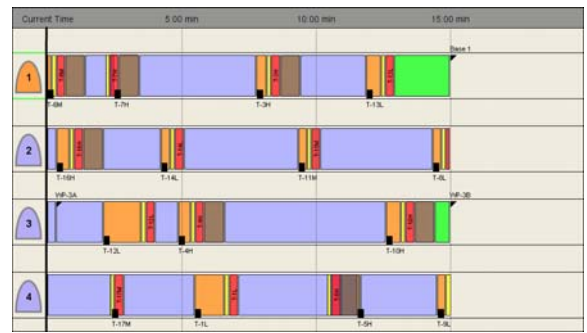
PAGE 1 OF 2

~~SECRET~~
 UNCLASSIFIED

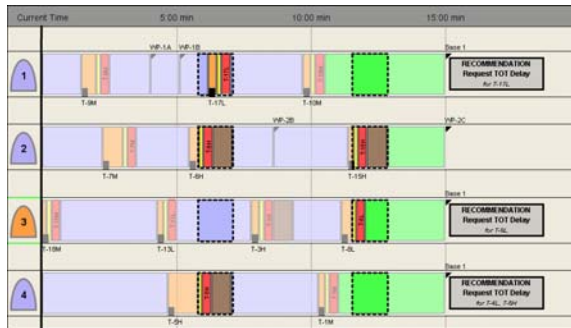
Figure 3: Air Tasking Order Example

Task	Current Target TOI / EIA	Next Waypoint or Loiterpoint	Next Expected Action	Upcoming Active Targets TOI / EIA			Mission Finish
1	T-6M 12:05:50Z - 12:06:15Z 12:05:34Z	-	Arming for T-6M 12:05:45 - 12:05:55	T-3H 12:13:55Z - 12:14:15Z 12:13:22Z	T-12K 12:18:50Z - 12:19:30Z 12:17:37Z	-	Base 1 12:23:30Z
2	T-10H 12:06:35Z - 12:06:55Z 12:05:50Z	-	Arming for T-10H 12:06:25 - 12:06:55	T-14L 12:10:20Z - 12:10:40Z 12:09:48Z	T-11M 12:11:50Z - 12:11:50Z 12:14:50Z	T-8L 12:20:25Z - 12:20:45Z 12:19:56Z	Base 1 12:23:00Z
3	T-12L 12:09:15Z - 12:09:35Z 12:07:41Z	WP-3A 12:09:48Z	Arming for T-12L 12:09:05 - 12:09:35	T-4H 12:11:00Z - 12:11:20Z 12:10:28Z	T-5H 12:18:50Z - 12:19:15Z 12:18:11Z	-	Base 1 12:26:19Z
4	T-17M 12:08:55Z - 12:09:25Z 12:07:53Z	WP-4A 12:06:48Z	Arming for T-17M 12:07:55 - 12:08:25	T-8L 12:18:55Z - 12:11:15Z 12:11:04Z	T-5H 12:18:15Z - 12:18:30Z 12:18:42Z	T-8L 12:20:35Z - 12:20:55Z 12:20:58Z	Base 1 12:25:18Z

(a) Manual



(b) Passive



(c) Active



(d) Super Active

Figure 4: The Four Possible Levels of Automated Decision Support in MAUVE

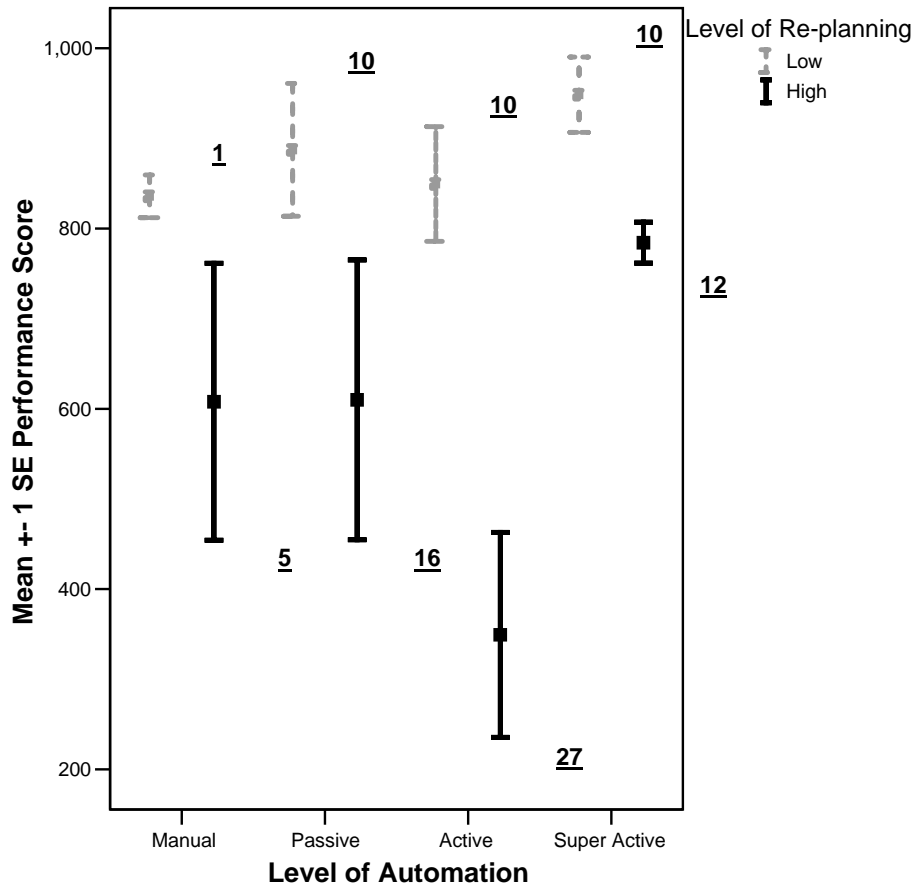


Figure 5: Performance Score and TOT Request Results

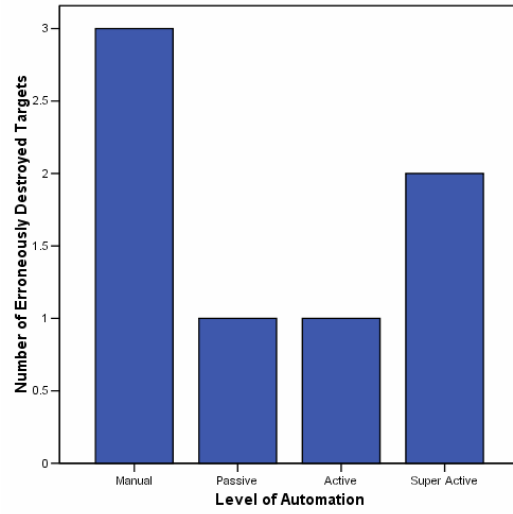
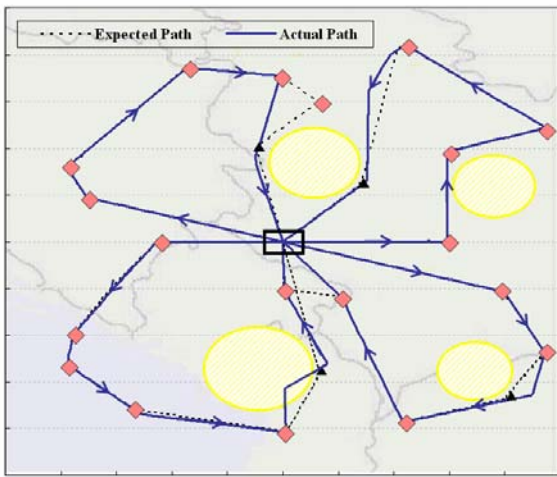
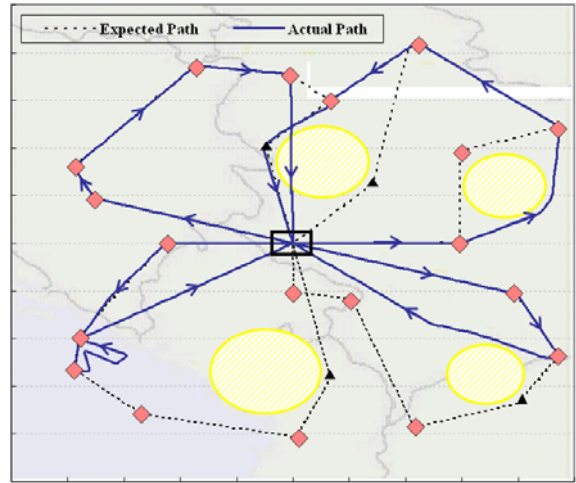


Figure 6: Errors of Commission across Levels of Automation



(a)



(b)

Figure 7: Cognitive Shedding Strategies for Operators under High Workload