Managing Multiple UAVs through a Timeline Display

M.L. Cummings[*] and P.J. Mitchell.[†]
Massachusetts Institute of Technology, Cambridge MA 02139

Network-centric operations, in which both automated and human entities are linked in order to leverage information superiority, will bring increases in available information sources, volume of information and operational tempo, placing higher cognitive demands on operators. 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 compliment operators, and how human cognitive limitations will impact overall system effectiveness. To this end, this paper presents the results of an experiment in which an operator simultaneously managed four highly autonomous UAVs executing an air tasking order, with the overall goal of destroying a predetermined set of targets within a limited time period. The primary factors under investigation were different levels of automation from manual to management-by-exception represented in a timeline, as well as different levels of replanning which are needed in time-sensitive targeting scenarios. Increasing levels of automation can reduce workload but they can also result in situation awareness degradation as well as automation bias. Results demonstrate that operators became fixated on the need to globally optimize their schedules and did not adequately weigh uncertainty in their decisions. This fixation significantly degraded operator performance to the point that operators without any decision support performed better than those with probabilistic prediction information and the ability to negotiate potential outcomes.

I. Introduction

Human supervisory control (HSC) occurs when a human operator monitors a complex system and intermittently executes some level of control on the system though some automated agent. HSC tasks are primarily cognitive in nature and generally do not require constant attention and/or control. For example, a single air traffic controller can handle multiple aircraft because the onboard pilots handle the flying task, while the controller is primarily concerned with navigation tasks that do not require constant attention. Similarly, while many operators are presently needed to control a single unmanned aerial vehicle (UAV), as technology and autonomous control improve, automation will handle the task of flying, thus enabling the individual controller 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 compliment 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 command and control, efficiently allocating attention between a set of dynamic tasks becomes critical to both human and system performance. However, managing multiple vehicles increases the number of available information sources, volume of information and operational tempo, all which place higher cognitive demands on operators. 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 compliment operators, and how human cognitive limitations will impact overall system effectiveness. To this end, this paper will discuss recent efforts to model human supervisory control of multiple vehicles in terms of increasing levels of automated decision support and the impact on operator control strategies.

[*] Assistant Professor, Aeronautics & Astronautics, 77 Massachusetts Ave, 33-305, Senior Member
[†] Research Scientist, Aeronautics & Astronautics, 77 Massachusetts Ave, 33-309

II. Embedded Human Supervisory Control Loops

The role of automated decision support in supervisory control for multiple UAVs has been studied by Ruff et al.[2] and Wickens et al.[3], but both of these studies assumed that UAV operators would have some degree of responsibility for actually flying the UAVs in addition to other tasks such as navigation and payload management. Both of these studies demonstrate that because UAV operators are assigned a number of control responsibilities which include flying, navigation, and mission execution, their ability to control multiple vehicles is limited, even with automated decision support. While not explicitly addressed in these studies, they highlight that human supervisory control for management of one or more UAVs is a nested control problem as represented in Figure 1.

The inner loop of Figure 1 represents a basic guidance and motion control loop which is the most critical loop that must obey aerodynamic constraints or the UAV will crash. 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. Lastly, the final 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. 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. If humans must 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. Violations of the priority scheme represented in Figure 1 have led to serious problems exemplified by the numerous Predators crashes. When operators become cognitively saturated or do not correctly allocate their cognitive resources to the appropriate control loops in the correct priorities, they violate the control loops constraints, potentially causing catastrophic failure.

In terms of operator cognitive interactions, the three loops in Figure 1 are represented by the hierarchical skill, rule, and knowledge-based behavior taxonomy[4]. Flying an aircraft is a skill-based behavior (SBB) because it requires primarily human input driven by sensory perception and automaticity which occurs with significant training. In UAV operations, rule-based behaviors (RBB) are required for procedural actions driven by underlying if-then-else algorithms, which is representative of navigation tasks. For example, if a UAV suffers a system failure and must return to base, UAV operators execute established procedures to redirect the aircraft home, avoiding all obstacles. Lastly, the highest level of cognitive control, knowledge-based behaviors (KBB), which require situation assessment, evaluation, and judgment, all are needed for mission and payload management.

Generally automation is effective in reducing the workload associated with skill-based behaviors. For example, autopilot altitude hold significantly reduces pilot workload because it automates a skill-based function and frees cognitive resources for higher level rule and knowledge-based behaviors. In terms of the navigation control loop, some rule-based behaviors can be automated such as path planners that ensure obstacles and no-fly zones are avoided, however, automation of navigation RBB is limited in that algorithms must be comprehensive and reliable. Because command and control navigation is inherently uncertain with a large number of known and unknown variables and constraints, this level is difficult to effectively automate. Lastly, because KBBs require judgment, intuition, and naturalistic decision-making, introducing automation is much more difficult at this level.

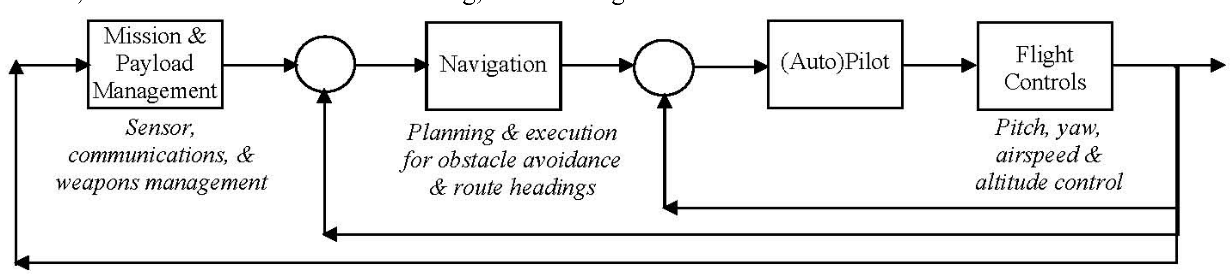


Figure 1: Unmanned Air Vehicle Supervisory Control Loops

A. Levels of Automation

The challenge in achieving the one-controlling-many goal for management of multiple unmanned vehicles in the future is not only to determine if automation can be used to reduce workload, but to what degree in each of the control loops in Figure 1. Automation strategies can range from fully automatic where the operator is completely left out of the decision and control process to minimal levels where the automation offers basic data filtering or recommendations for the human to consider. Sheridan and Verplank[1] outlined a scale from 1-10 where each level represents progressively more automation for decision and action selection, as shown in Table 1. For SBBs, higher levels of automation (LOAs) in general result in lower workload. However, RBBS and to a greater extent, KBBs require lower levels of automation because highly and even partially automated systems can result in measurable costs in human performance, such as loss of situational awareness, complacency, skill degradation, and decision biases [5, 6].

A few studies have investigated levels of automation in the context of multiple UAV supervisory performance. Ruff et al.[2, 7] 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 one study, the moderate LOA produced the highest levels of operator situation awareness (SA) and performance, however in a subsequent study there was no difference.

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.

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 strategies or decision support perspective. Wickens et al. [3, 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 managements control loops was helpful in reducing operator workload. However, they also did not distinguish between levels of automation between the RBB navigation and KBB mission management loops thus it is not clear how one level of automation in one loop affected both the other and the outcome of the mission.

In order to address this gap in understanding how LOAs in a single loop affect human and system performance, we developed a UAV interface with four different levels of automation to specifically address how increasing automation strategies would affect the KBB mission and payload management loop. In our study, the guidance and motion control loop was fully automated such that pilots did not have to intervene to control any flight axis (Level 10). The navigation LOA was held constant, so the UAVs' heading control was fully automated (Level 10). The KBB mission and payload management loop, the most difficult to automate, was the only loop that varied in automation level. We isolated the mission and payload management loop from the guidance and motion, and navigation control loops so that the results would not be confounded with possible interference and interaction from the two lower-level nested cognitive control loops.

III. MAUVE: The Experimental Test Interface

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 dual screen simulation test bed named the Multi-Aerial Unmanned Vehicle Experiment (MAUVE) interface was developed (Figure 2). This interface allows an operator to effectively supervise four UAVs simultaneously, and intervene as the situation requires it. In this simulation, users take 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. As discussed previously, the guidance and motion control loop was fully automated as was the basic navigation control loop. Because the simulated UAVs were highly autonomous, they only required that operators provide high level mission planning and execution actions as inputs to the UAVs. The UAVs launched with a pre-determined mission plan that came from an air tasking order (ATO), so initial target assignments and routes were already completed. The operator's job in the MAUVE simulation was to monitor each UAV's progress, re-plan aspects of the mission in reaction to unexpected events (a KBB), and in some cases manually execute mission critical actions such as arming and firing of payloads. As will be demonstrated in later sections, arming and firing, while arguably a RBB, was a definite KBB when contingencies arose.

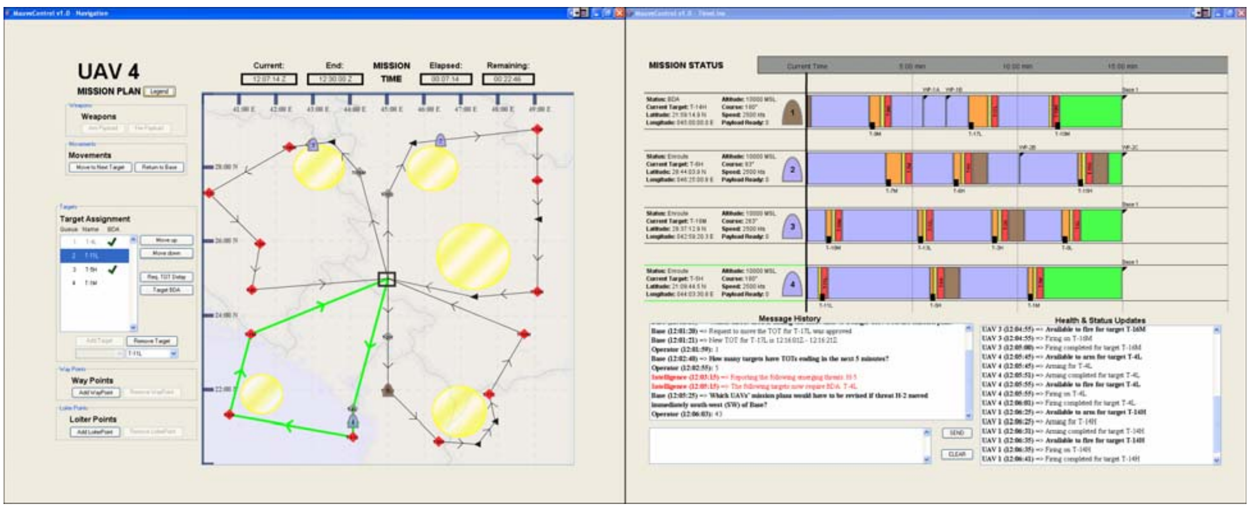


Figure 2: The MAUVE dual screen interface

The UAVs supervised by subjects in MAUVE were capable of 6 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 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. Performing BDA was not required for every target, but was dependent on preplanning or in-flight contingencies.

A. Navigation Display

The left-hand side of the MAUVE interface is known as the navigation display, and it consists of a mission time window, map display, and a mission planning and execution bar (Figure 2, left side). A large mission time box showing both time elapsed and time remaining in absolute and relative terms is located on the top right of the map display. The map display represents a two-dimensional spatial layout of the battlespace, updated in real-time. Threat or hazard areas, circular in shape, have a striped yellow coloring pattern, and can be dynamic throughout scenarios, changing size, locations, disappearing entirely, or emerging as time progresses. The UAVs, always held constant at four, independently change colors according to their current action (Table 2). The thick light green line around one of the mission plans indicates that plan is currently selected by

Table 2: Color-Coded UAV Stages

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

the user.

Targets are designated by a diamond-shaped icon, and are assigned a relative importance to the mission plan (priority) of high (H), medium (M), or low (L). Active targets are differentiated from inactive targets by their color, which is either red or gray on the display, respectively. An inactive target is any target that has either already been destroyed, or its TOT deadline missed. Waypoints, shown on the map display with black triangle icons, are UAVs turning points enroute. Functionally, a loiter point is similar to a waypoint except that when a UAV reaches a loiter point, the UAV can loiter for a user-specified amount of time before moving on. UAV routes on the map display can be changed in minor ways by selecting a particular waypoint or loiter point and dragging it to the desired location. More major 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. Routing changes were only required as a result of unexpected scenarios and represents real-time replanning.

Operators are provided with a "Request TOT Delay" button which allows them limited opportunities to manipulate the time-on-targets (TOTs) for those targets assigned. 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 that target and therefore will miss their deadline, or 2) for workload purposes, i.e., if an operator feels they need to spread out a very high workload time block to manage the UAVs more effectively. However, this function should be used with care because moving back one target's deadline likely affects the UAV's arrival time at all subsequent targets. It is important to highlight this change of TOT is a request, not a command, and operators' requests can be approved or denied. The probability of approval is a function of how far in advance of the deadline the request is sent, as would likely be the case in true military situations. The probability distribution for approval is given by Equation 1, which is not known by operators.

$$P(Approval) = 1.0 - e^{-t/450}, \text{ t = time of request in seconds before deadline} \quad (1)$$

However they were briefed that when a TOT deadline is immediately approaching, the chance of approval is zero, but nearly 1.0 when requested 15 minutes in advance, which is the limit for the decision support, which will be discussed in the next section. A request always takes 5 seconds for response, and during this intervening time no other TOT requests can be made. Users can request as many TOT delays as they wished for a given target, but there are no guarantees of approval.

Command buttons for the UAVs include Arm Payload, Fire Payload, Move to Next Target, and Return to Base. Arming and firing are only enabled if the pre-established rules of engagement (RoE) of the simulation are met. For arming, the UAV must be directly on top of a target within the arming or firing windows, and for firing, the UAV should be armed at the correct target. The Move to Next Target button allows operators to bring UAVs out of loiter points in case of scheduling problems and the Return to Base button causes all future targets, waypoints and loiter points to be deleted from the mission plan. Subsequently a straight line path is planned directly back to base and is intended for off-nominal scenarios.

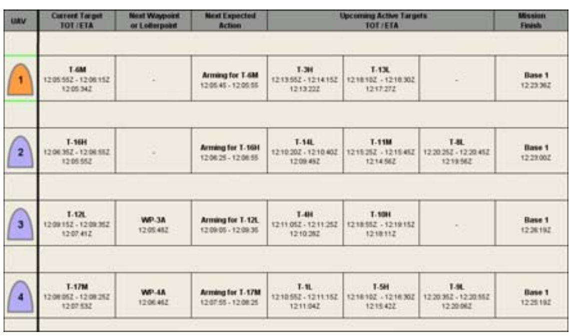
B. Decision Support Display

The right-hand side of the MAUVE simulation in Figure 2 provides decision support, and it consists of a UAV status window, chat box, UAV health and status updates, and the decision support window. The status window at the top left of the decision support display gives operators low level, detailed information for each UAV such as current target, current action being performed, position in latitude and longitude, course, and weapons information. Speed and altitude are also shown in the status display, although they are not directly controllable by operators.

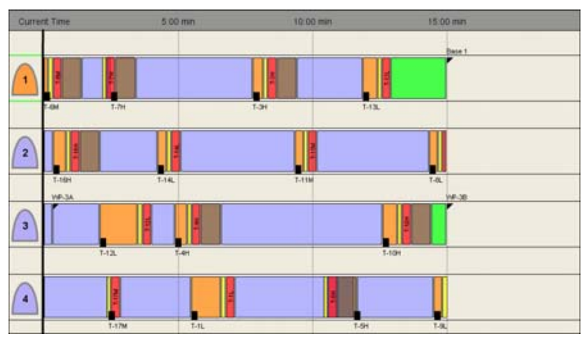
The bottom left of the decision support display (right side, Figure 2) has a text-based communication tool known as a chat box that contains a time history of all human communication interactions. The chat box is included because it is an established method of communications in current day military command and control scenarios[9], and is an externally valid embedded secondary workload and situation awareness measurement tool[10]. The chat box window displays various notification messages that appear in response to scenario events or actions taken by users, as well as periodic task-relevant questions for operators to answer. The accuracy and time delay in responses to the online queries from a confederate automated superior can be measured to obtain an objective measurement of situation awareness as well as secondary workload, or spare capacity. 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.

The decision support always appears in the top right of the decision support display and the manipulation of the appearance and functionality of this window is the primary independent variable for the experiment that will be discussed in the subsequent section. The basic premise of the decision support is to simplify standard air tasking order (ATO) data and combine it in a single interface with up-to-date mission planning information. An ATO provides a schedule of events and required resources needed over a period of hours and/or days. Information contained in an ATO includes which aircraft have been assigned to certain strikes, times on targets, way points that must be flown on those strikes, and call signs to be used on those missions. ATOs are complex and hard to interpret, particularly under time pressure. Despite this, operators are still expected to extract the information they need in a timely manner. While some level of decision support is required to more effectively manage ATO information and scheduling, it is not clear what level of automation will provide 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 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 are four possible forms of decision support in MAUVE that roughly correspond to levels 1, 2, 4, and 6 (Table 1), termed manual, passive, active, and super active respectively.

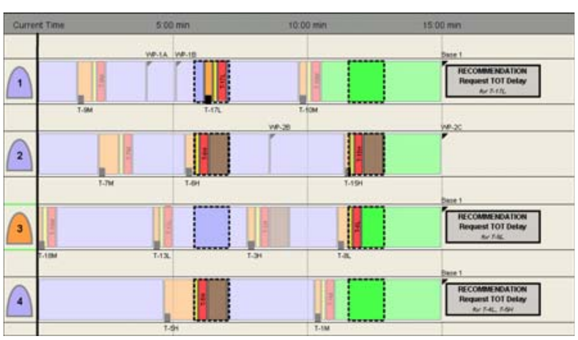
The manual LOA level of decision support (Figure 3a) presents all required ATO and mission planning information in a text-based table format. Under the “Current Target” and “Upcoming Active Targets” headings, current TOT windows and ETAs for up to the next 4 targets in queue are presented for easy comparison. ETAs for arrival at base in the “Mission Finish” column and the next waypoint or navigation point on the current route segment (if applicable) under “Next Waypoint or Loiterpoint” are also given. Further assistance is provided to the user through the “Next Expected Action” column, which told the user what they should be doing next and at what time, according to the ATO. This information is updated dynamically to reflect changing ATO requirements and mission planning changes initiated by the user.



(a) Manual



(b) Passive



(c) Active

(d) Super Active

Figure 3: The four possible levels of decision support in MAUVE

The passive LOA (Figure 3b) 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 (Table 2). The major difference between the passive and the manual level is entire schedule integration for users instead of them having to piece it all together, as illustrated in Figure 4. The visual timelines are relative, with the left side representing predicted UAV actions in the near future and the right side up to 15 minutes into the future. Figure 4 illustrates the standard elements of a representative visual timeline. Target ETAs are represented by black rectangles on the bottom of each timeline, and waypoint, loiter point and base arrival times are marked by black triangles on the top of each timeline. The 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.

With this visual representation, recognizing problems with the current schedule is perceptually-based, allowing users to visually compare the relative location of display elements instead of specific times to one another. Another emergent feature of this display is that if the color of the UAV icon to the immediate left of the timeline is not the same color as the icon on the left side of the timeline, the operator can immediately recognize that the UAV needs attention. This level of decision support is termed passive because the automation is not performing any tasks except transforming the basic ATO and mission planning information into graphical format.

The active LOA (Figure 3c) uses the same horizontal timeline format as the passive automation level, but provides additional help from the computer. In the active version, an algorithm searches for periods of time in the schedule that it predicts will cause high workload for the operator, directing the operator's attention towards them. 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. 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. As no information is hidden, only made less salient, the operator's attention can be directed to the appropriate areas of the schedule while allowing them to maintain SA for the rest of the mission.

In addition to identifying areas of high workload, the computer also recommends a course of action to alleviate the high workload areas, such as moving a particular TOT. Tulga and Sheridan[11] demonstrated that even with preview, the time subjects plan ahead typically decreases as workload increases, so we hypothesized that by providing both preview into the future as well automated recommendations, the need for planning time would be decreased and thus performance would improve. To this end, computer recommendations appear in gray boxes to the right of each relevant UAV's timeline and subjects have several options: 1) They can acknowledge a high workload area but take no action, 2) They can follow the recommendation to relieve the projected high workload area by shifting a TOT, or 3) They can make other mission planning changes to ensure the high workload area does not occur, such as un-assigning a target from a UAV's mission plan. While the automation made locally optimal recommendations, the algorithm was not globally optimal. Following the computer's recommendation to relieve a high workload area always removed that particular schedule conflict, but sometimes created another in the process.

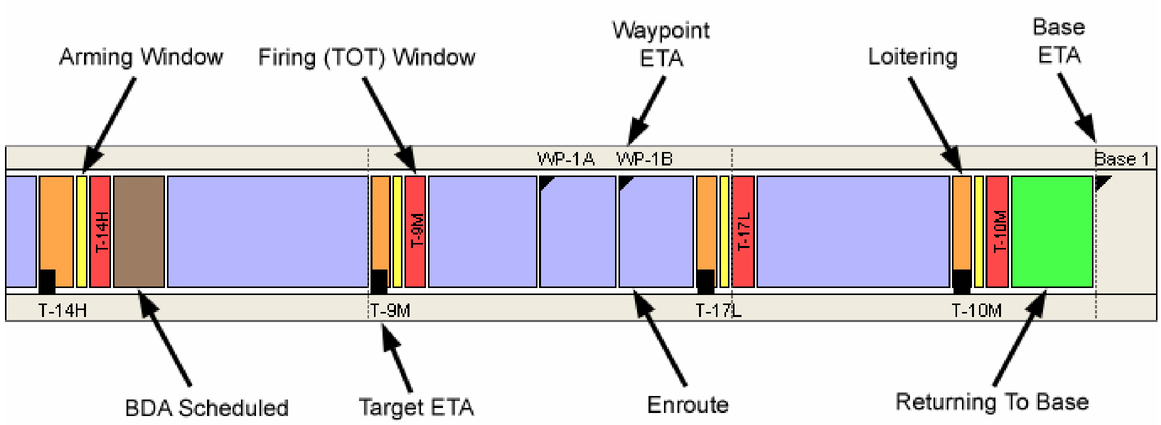


Figure 4: A representative visual timeline

The reverse shading technique in conjunction with the recommendations permits operators to make local changes to alleviate workload and immediately see the effect on the global mission plans of all the UAVs without requiring any drilldown to subsequent screens to solve the problem. The purpose of this level of automation was to help operators to identify time periods of potential concern where they could be overwhelmed farther in advance, so that they could better plan to avoid them, or at least be better prepared to handle them. This level of decision support was termed active because the automation narrowed down a set of possible solution alternatives for high workload problems to a single recommendation.

The super active LOA (Figure 3d) also builds upon the passive level visual timeline, but instead of making recommendations to the operator as in the active LOA, a management-by-exception approach is taken whereby the computer automatically executes the arming and firing actions for all UAVs at each target, when the rules of engagement for such actions are met. For example, in order to fire, a UAV has to be located at the particular target it was due to fire on, already armed, and the scenario time within the TOT window for that target. While the automation handles the actual execution of tasks, 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 arrives at the correct targets on time. Up to 30 seconds in advance before every arming and firing action, exception boxes appeared to the right of the timeline that allowed the operator to veto these actions. The color of the box indicated which action the UAV was preparing to perform: red for firing and yellow for arming. This level of decision support was termed super active because the automation was performing all of the mission critical execution actions for the user.

IV. The Experiment

In order to investigate human knowledge-based behaviors in time management and 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 an increase in workload would affect performance and operator situational awareness.

A. Apparatus, Participants, and Procedure

Training and testing was conducted on a four screen system called the multi-modal workstation (MMWS) [12], originally designed by the Space and Naval Warfare (SPAWAR) Systems Center as a test prototype to aid the development of human-computer interface recommendations for future Navy command and control systems. The top three screens used were 21" and were run at 1280 x 1024 pixels, 16-bit color resolution, while the bottom screen was 15" and was run at 1024 x 768 pixels, 32-bit color resolution. The workstation was a Dell Optiplex GX280 with a Pentium 4 processor and an Appian Jeronimo Pro 4-Port graphics card. Subjects interacted with the simulation via a Logitech MX500 cordless mouse and a generic numeric key pad. During testing, all mouse clicks and both message box histories, including incoming and outgoing messages, were recorded by software. In addition, screenshots of both simulation screens were taken approximately every two minutes, all four UAV locations were recorded every 10 seconds, and whenever a UAV's status changed, the time and change made were noted in the data file.

A total of 12 subjects took part in this experiment, 10 men and 2 women. Subjects were recruited based on whether they had UAV, military and/or pilot experience. The subject population consisted of a combination of students, both undergraduates and graduates, as well as those from the local reserve officer training corps (ROTC) and active duty military personnel. All were paid $10 on 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. Nine participants were members of the ROTC or active duty USAF officers, including seven 2nd Lieutenants, a Major and a Lieutenant Colonel. While no subjects had large-scale UAV experience, 9 participants had piloting experience. The average number of flight hours among this group was 120.

Subjects had two main objectives in this experiment: 1) To guide each UAV's actions so that together, all UAVs under their supervision properly executed the required missions and engagements of the current ATO, which changed over time, and 2) To answer periodic questions about the situation from commanders. All subjects 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. The experimenter observed and took notes on how subjects responded to re-planning events answered questions. Each simulation was run several times faster than real time so an entire strike could take place over 30 minutes (instead of several hours as is commonplace in real life strikes). Figure 5 depicts the experimental test bed with the center top and top right screens contained the MAUVE simulation navigation and decision support windows, respectively. The top left screen contained subjects' objectives in rank priority order for the scenarios, and was static throughout the experiment, while the bottom screen contained the color coding for UAV actions in the simulation (Table 2). This information was provided during testing because of pre-test feedback that indicated it was a useful reminder and to ensure that participants were always aware of the rules of engagement.

Figure 5: The MMWS Running MAUVE

B. Experimental Design

Two independent variables were of interest in this experiment: level of decision support (Figure 3) and level of re-planning. For re-planning, operators were required to handle unplanned contingencies such as emergent threat areas and targets, new tasking from superiors to either add or delete BDA from a mission, as well as deal with any system failures that might require a UAV to return to base unexpectedly. For the replanning experimental factor, low and high levels of schedule re-planning were investigated. The low re-planning condition contained 7 re-planning events, while the high re-planning condition contained 13. Groups of re-planning events were interspersed at approximately 3 minute intervals, but under the low re-planning level these groups only consisted of a single event. Under high re-planning, some groups were composed of 2 or 3 re-plans occurring within 60 seconds of each other. The level of decision support was a between subjects variable and the level of re-planning was a within-subjects repeated variable, so subjects were randomly assigned to a LOA factor level but saw all re-planning conditions.

The dependent variables were a performance score and situation awareness scores. The performance score measured how well subjects 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. A score of zero on this rating scale indicated no objectives had been met in the session, while a perfect score of 1000 meant that all mission objectives were met.

Situation awareness (SA) is generally defined as the perception of the elements in the environment within a volume of time and space, and the comprehension of their meaning and the projection of their status in the near future[13]. SA has three distinct levels, which are: 1) the perception of the elements in the environment, 2) the comprehension of the current situation, and 3) the projection of future status[13, 14]. Situation awareness was captured through a subjective SA scale constructed from expert observer ratings to test the validity of the objective SA measure. Subjective SA scales based upon expert observer ratings have been found to be a reliable and valid measure way to measure SA[15]. This scale was based upon expert opinion on what were judged to be the four most salient indicators of SA in the test sessions:

1. Number of large scale threat area incursions where there was no operator response to re-plan that UAV's route. A large scale threat incursion was defined as when a UAV received 3 or more hits from a threat area before exiting.
2. Amount of system wait time at targets due to loss of situation awareness.
3. Number of targets missed due to lack of situation awareness.
4. Percentage of re-planning events successfully completed.

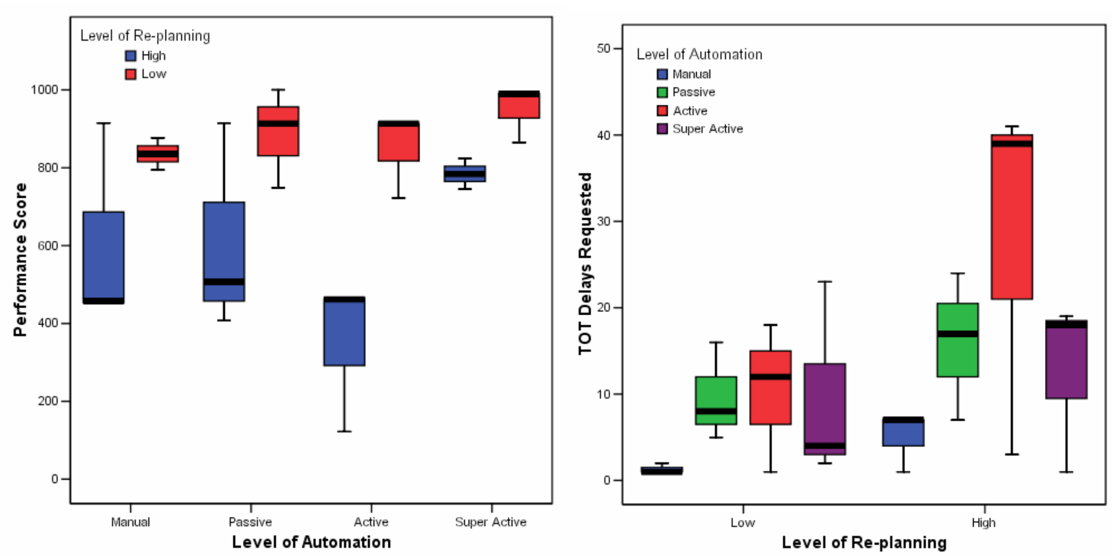
The first two indicators were measures of level 2 SA, while the second two were measures of level 3 SA. Each test session was assigned four individual scores, one for each indicator, and a score for Level 2 SA was obtained by averaging the ratings for indicators 1 and 2, while a score for Level 3 SA was obtained similarly with indicators 3 and 4.

V. Results

For the statistical analysis of the performance scores, a 2x4(3) repeated measures linear mixed model was used. Three subjects were nested within each of the four automation levels, and the two levels of re-planning (within-subjects) were low and high. Age was used as a covariate in the analysis, and for all reported results $\alpha = 0.05$ unless stated otherwise. As previously stated, the performance score was an aggregate measure of overall human performance that incorporated the number and priority of targets correctly destroyed, number of times BDA was correctly performed, and any penalties from threat hits, late arrivals to base, or targets erroneously destroyed. Figure 5a shows the average performance scores for each experimental condition. Level of re-planning was significant ($F(1,9.9) = 19.40$, $p = 0.001$) while level of automation was marginally significant ($F(3,10.8) = 3.04$, $p = 0.076$). There was no significant interaction between the factors.

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. The significant driver of this problem was found to be operators' use of the "TOT Delay" function. Through requesting TOT delays, operators could manipulate target deadlines to spread out their workload and/or make deadlines they otherwise would have missed due to inadequate prior planning and/or execution. The count of TOT delays requested was examined using non-parametric analysis and a Kruskal-Wallis test showed that the number of TOT delays requested was marginally significant between automation levels ($p = 0.096$). Figure 5b reveals that the difference in number of TOT delay requests was driven by a lower number for the manual level of automation across both levels of re-planning (Wilcoxon Rank Sum Test, $p = 0.015$), and a higher number of TOT requests in the active level, predominantly for high replanning ($p = 0.065$).

A major consideration in the evaluation of all HSC systems is situation awareness. For level 2 SA measures indicating how well subjects comprehended the current situation, a Wilcoxon Signed Rank Test 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 ($p = 0.084$). Operators using the super active level had higher level 2 SA. As depicted in Figure 6, subjects with the active level of automation had the lowest level 2 SA. For level 3 SA measures that captured subjects' ability to project the current situation into the future, a Wilcoxon Signed Rank test demonstrated that subjects' level 3 SA for the high re-planning scenario was significantly lower ($p = 0.004$) than for the low workload condition, but there was no effect of level of automation.



(a) **(b)**

Figure 5: Performance Score & TOT Request Results

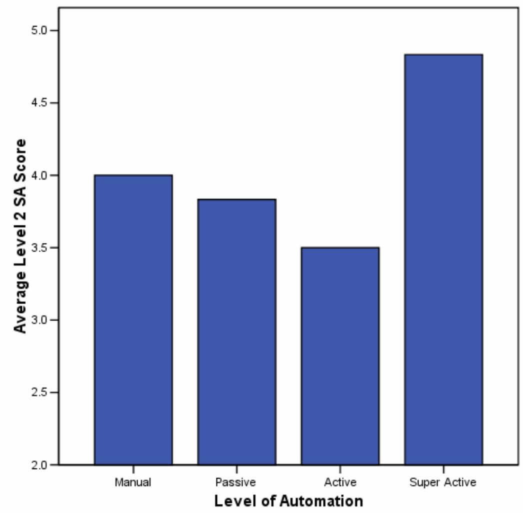


Figure 6: Level 2 SA Results

VI. Discussion

The significant difference in performance scores across re-planning levels in this experiment indicates that the number of re-planning events is an important influence on command and control mission difficulty, as expected. Under low levels of replanning, subjects at all levels of automation performed equally well, but under high re-planning, subjects using active automation had lower performance scores than all other automation levels. The poor overall performance of the active level, indicated by both poor performance scores and poor SA of the subjects, was unexpected. It was expected that the automated decision support would allow subjects to plan ahead more efficiently as demonstrated by Ruff et al.[2] who found that a medium level of automation called management-by-consent, corresponding most closely to the active level in this study, had performance advantages over manual and management-by-exception (super active) conditions for multiple UAV supervision. Many previous studies from other domains have also found collaborative types of automation to have superior performance to highly automated or manual systems[16-18]. However, a primary limitation in complex systems as demonstrated in this study is that users can often be overwhelmed by the large array of possible actions. This can be particularly problematic for command and control systems under temporal constraints and with significant uncertainty, as was the case in this study.

The performance decrement under the active level of automation can be attributed to subjects' inappropriate use of the "Request TOT change" function in the MAUVE simulation. The ability to move a target's TOT was an intervention meant to be used sparingly, and the consequences of doing so were not often adequately considered by operators. It is important to highlight that all subjects with all levels of automation had the ability to request a TOT change. However, only the subjects with the active automation that predicted future areas of potential high workload used it to excess and thus were significantly negatively impacted. The question must then be asked, why is it that such a simple graphical shading technique combined with automated recommendations (a relatively small change compared to the passive level visualization), had such a dramatic impact on the use of the Request TOT change" function.

A. Global Optimization and Stopping Rule Generation

In the active level of automation, subjects clearly were unable to generate appropriate stopping rules when trying to achieve a particular schedule move. Stopping rules are the criteria that individuals use to "satisfice" in uncertain situations, i.e. choosing the current best plan that is good enough. While often humans can adapt effective heuristics for generating stopping rules [19], it is particularly difficult for them to do under dynamic, nonstationary processes typical of command and control domains[20, 21]. Moray et al.[22] found that even if subjects were given an optimal scheduling rule, they were unable to implement it under enough time pressure, resorting instead to significantly non-optimal heuristic rules. In the context of this experiment, subjects could not effectively decide when to stop attempting to optimize (minimize) one or more future workload bottlenecks, even though the rules were straightforward and they were warned not to overuse the TOT change request. At the detriment of other tasks and vehicles requiring their attention, subjects often focused on obtaining a particular delay until they obtained it. In seeking to globally optimize their schedules, subjects were actually narrowly focused on the timeline display, and because this global optimization task was a very high workload process, the operators became overloaded and performed poorly.

Moreover, there were clear indications that operators were not effectively integrating probabilistic information that would guide these stopping rules. In the active level of automation, subjects appeared to be unable to effectively integrate the probabilistic nature of future workload predictions in that the automation only highlighted possible future areas of high workload and the farther into the future, the less likely a potential bottleneck. However, despite prior training to the contrary, a low probability predicted area of high workload 15 minutes into the future was generally treated the same as high probability workload saturation 1 minute away. Moreover, when subjects realized they had a near-term overload situation occurring, rather than cutting their losses and choosing to give up on a target to improve a UAV's arrival time at subsequent targets, subjects often tried until the very last possible instant to obtain the TOT delay for the UAV with immediate scheduling difficulties. As discussed previously, the longer subjects waited to request a TOT change, the less likely they were to receive an approval. Despite this fact, it appeared that operators believed they could always make up for lost time through the request button, when in reality the probability of approval became unreasonably low, often zero. This behavior was likely due to erroneous judgment of the probability of obtaining a last minute approval, as humans are not good estimators of chance and typically overestimate very small probabilities[23].

B. Preview Visualization

In this study, it is not clear how the problems with stopping rule generation and probabilistic inference were exacerbated by potential problems with the shading preview visualization technique. Past research investigating preview is mixed in regards to its usefulness. While preview in manual control of a moving entity such as piloting a plane is a well-known and universally integrated decision tool in cockpits, there is significantly less research addressing the usefulness of preview in scheduling and managing tasks, especially those under time pressure. Preview has been shown to lead to erroneous heuristics[24], while another study found aiding pilots in attending to lower priority secondary tasks through the use of workload preview had no impact on performance[25]. However, task scheduling results from a process control study suggest the lack of effective preview could be from difficulties in display interpretation[26], which could also be a factor in this present study. Current research with this work is investigating this potential problem source and possible design strategies, both graphically and algorithm-based, to improve performance and mitigate bias.

VII. Conclusion

Despite its many benefits, intelligent autonomy cannot replace the human in the knowledge-based role in the command and control mission management loop. Automation should support human knowledge-based decision making while also mitigating automation bias[27], which is particularly problematic in command and control domains[28]. To this end, a study was conducted to investigate how levels of automation affect the UAV knowledge-based mission and payload management control loop from a human supervisory control perspective. The goal was to determine how increasing levels of automation affected operator performance in order to identify possible future automation strategies for multiple independent UAV scheduling and management. A human-in-the-loop experiment revealed that when provided with a high workload preview visualization as well as automated recommendations for workload mitigation, operators became fixated on the need to globally optimize their schedules, and did not adequately weigh uncertainty in their decisions. This fixation prevented subjects from generating effective stopping rules and significantly degraded their performance to the point that operators without any decision support performed better than those with automated decision support.

This research motivates the need to develop robust scheduling decision aids that convey both possible options and uncertainty, while also effectively bounding the operator such that satisficing occurs in a time-constrained environment. We term this approach as "bounded collaboration" since a robust decision tool is needed that allows humans use of their judgment, experience, and pattern recognition strengths but also constrains users so they do not revert to biased and potentially catastrophic heuristics. Future work with this testbed will include investigating to what degree workload prediction visualization helps or hinders operator performance as well as how to incorporate more sophisticated scheduling algorithms to improve both human and system performance.

Acknowledgments

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

References

[1] Sheridan, T. B.; Verplank, W. L., "Human and Computer Control of Undersea Teleoperators", Man-Machine Systems Laboratory Report, MIT, Cambridge, 1978.

[2] Ruff, H. A.; Narayanan, S.; Draper, M. H., Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles, *Presence*, 11, 2002, 335-351.

[3] Wickens, C. D.; Dixon, S.; Chang, D., "Using Interference Models to Predict Performance in a Multiple-Task UAV Environment - 2 UAVs", AHFD-03-9/MAAD-03-1, Aviation Human Factors Division, Institute of Aviation, University of Illinois at Urbana-Champaign, Savoy, IL, 2003.

[4] Rasmussen, J., Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distractions in Human Performance Models, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, 1983, 257-266.

[5] Kaber, D. B.; Endsley, M. R.; Onal, E., Design of Automation for Telerobots and the Effect on Performance, Operator Situation Awareness and Subjective Workload, *Human Factors and Ergonomics in Manufacturing*, 10, 2000, 409-430.

[6] Parasuraman, R.; Sheridan, T. B.; Wickens, C. D., A Model for Types and Levels of Human Interaction with Automation, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30, 2000, 286-297.

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

[8] Wickens, C. D.; Dixon, S., "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] Cummings, M. L., The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations, *Adaptive Displays Conference*, Los Angeles, CA, 2004.

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

[11] Tulga, M. K.; Sheridan, T. B., Dynamic Decisions and Work Load in Multitask Supervisory Control, *IEEE Transactions on Systems, Man, and Cybernetics*, 10, 1980, 217-232.

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

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

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

[15] Bell, H. H.; Waag, W. L., Using observer ratings to assess situational awareness in tactical air environments, in Experimental analysis and measurement of situation awareness, Garland, D. J., Endsley, M. R. Eds., 1995,

[16] Vicente, K., *Cognitive Work Analysis. Toward safe, productive, and healthy computer-based work.*, Erlbaum, Mahwah, NJ, 1999.

[17] Layton, C.; Smith, P. J.; McCoy, E., Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation, *Human Factors*, 36, 1994, 94-119.

[18] Guerlain, S. A., Interactive advisory systems, *Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*, Savannah, Georgia, 2000.

[19] Gigerenzer, G.; Todd, P. M., *Simple Heuristics That Make Us Smart*, Oxford University Press, New York, 1999.

[20] Goonetilleke, R. S.; Drury, C. G., Human optimization with moving optima, *Ergonomics*, 32, 1989, 1207-1226.

[21] Klein, G., The Fiction of Optimization, in Bounded Rationality: The Adaptive Toolbox, Gigerenzer, G., Selton, R. Eds., MIT Press Cambridge, MA, 2001, 103-121.

[22] Moray, N.; Dessouky, M. I.; Kijowski, B. A., Strategic Behavior, Workload, and Performance in Task Scheduling, *Human Factors*, 33, 1991, 607-629.

[23] Tversky, A.; Kahneman, D., Judgment Under Uncertainty: Heuristics and Biases, *Science*, 185, 1974, 1124-1131.

[24] Wickens, C. D.; Pizarro, D.; Bell, B., Overconfidence, Preview, and Probability in Strategic Planning, *Humans Factors Society 35th Annual Meeting*, 1991.

[25] Segal, L. D.; Wickens, C., TASKILLAN II: pilot strategies for workload management, *Human Factors Society 34th Annual Meeting*, 1990.

[26] Sanderson, P., The human planning and scheduling role in advanced manufacturing systems: An emerging human factors domain, *Human Factors*, 31, 1989, 635-666.

[27] Skitka, L. J.; Mosier, K. L.; Burdick, M. D., Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 1999, 991-1006.

[28] Cummings, M. L., Automation Bias in Intelligent Time Critical Decision Support Systems, *AIAA Intelligent Systems*, Chicago, IL, 2004.