JOURNAL OF AEROSPACE COMPUTING, INFORMATION, AND COMMUNICATION
Vol. 3, June 2006

Automated Scheduling Decision Support for Supervisory Control of Multiple UAVs

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

In the future vision of allowing a single operator to control multiple unmanned vehicles (on land, in the air, or under water), it is not well understood how multiple vehicle control will affect operator workload, and what automated decision support strategies will improve, or possible degrade, operator performance. To this end, this paper presents the results of an experiment in which operators simultaneously managed four highly autonomous independent homogenous UAVs in a simulation, with the overall goal of destroying a predetermined set of targets within a limited time period. The primary factors under investigation were increasing levels of automation from manual to management-by-exception, manifested through a timeline visualization. Increasing levels of automation can reduce workload but they can also result in situation awareness degradation as well as complacency. This 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. These behaviors 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 workload 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

Received 21 August 2005; revision received 18 March 2006; accepted for publication 23 March 2006.

* Assistant Professor, Aeronautics & Astronautics, 77 Massachusetts Ave, 33-305, Senior Member
† Controls development engineer, Power Generation Technology, General Electric Energy

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 discusses recent efforts to investigate human supervisory control of multiple UAVs in terms of increasing levels of automated decision support and the impact of this decision support on workload mitigation strategies. More specifically, this paper focuses on how increasing automated assistance in schedule management of four independent homogeneous UAVs affected operator performance, workload, and situation awareness in an experimental setting. While previous studies have examined how well automated decision support affects operators' abilities to understand computer-generated plans,[2–4] this study extends previous research on operator strategies in temporal management of time constrained tasks[5] to examine how automated assistance affects these strategies and ultimately system performance in plan/schedule maintenance.

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.[6] and Wickens et al.,[7] 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 when 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 local directional 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 Predator 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.[8] 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

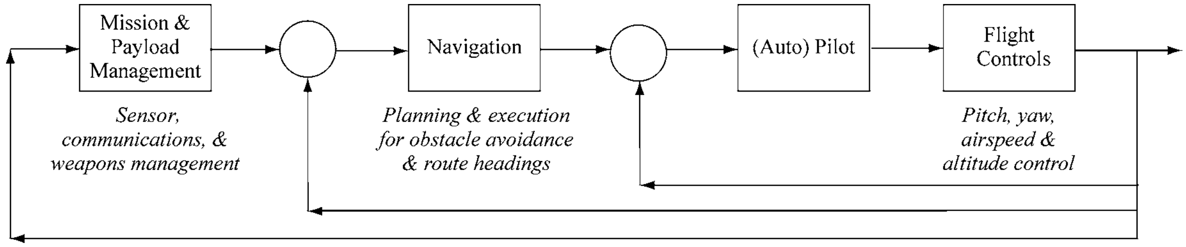


Fig. 1 Unmanned air vehicle supervisory control loops.

return to base, UAV operators execute established procedures to redirect the aircraft home, avoiding all obstacles. Lastly, the highest level of control, knowledge-based behaviors (KBB) require complex cognition such as situation assessment, evaluation, and judgment, which all are needed for UAV 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.

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.[9,10]

A few studies have investigated levels of automation in the context of multiple UAV supervisory performance. Ruff et al.[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 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.

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

Table 1 Levels of automation[‡].

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.

resources, either from an automation strategies or decision support perspective. Wickens et al. [7,11] 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 LOA impact on KBB loop human performance in schedule management, we developed a UAV interface with four different levels of automation to specifically address how increasing automation strategies would affect operator's abilities to manage a dynamic schedule. 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 UAV 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 control loops. The goal of this study was not to make any predictions about which LOA was better, as this would be task and context-sensitive in any real-world UAV application, but to determine how the different LOAs affected operator strategies when using a schedule management decision support tool.

III. MAUVE: The Experimental Test Interface

In order to study how levels of automation affect UAV HSC knowledge-based mission and payload management, specifically schedule maintenance, 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 supervise four UAVs simultaneously and intervene as the situation requires. 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 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, replan 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.

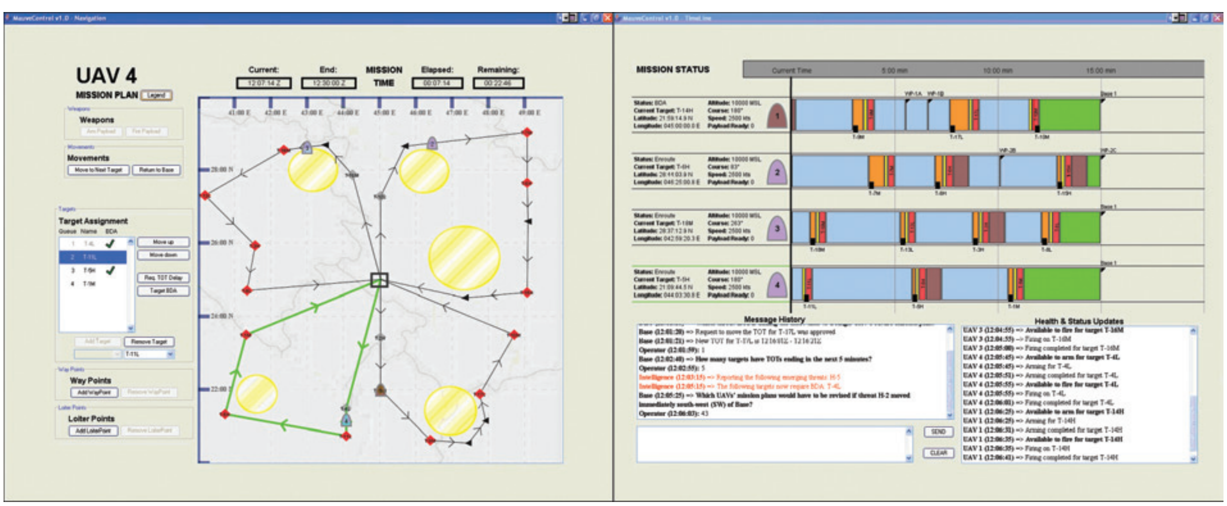


Fig. 2 The MAUVE dual screen interface.

The UAVs supervised by participants 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 weapons release where it is determined whether the weapon(s) hit the target and if the desired effect was achieved. In this simulation, 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 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). A thick light green line around one of the mission plans indicates that plan is currently selected by the user.

Targets are designated by a diamond-shaped icon, and are assigned a relative 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 UAV turning points. 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 to the next waypoint or target. 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 represent 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, participants were warned that 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 emphasize that 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.

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

When a TOT deadline is immediately approaching, the chance of approval is zero, but nearly 1.0 when requested 15 minutes in advance (participants were told this). A request always takes 2–5 seconds for response, and during this

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

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 patterns in case of scheduling problems. The Return to Base button causes all future targets, waypoints and loiter points to be deleted from the mission plan, and subsequently a straight line path is planned directly back to base.

B. Decision Support Display

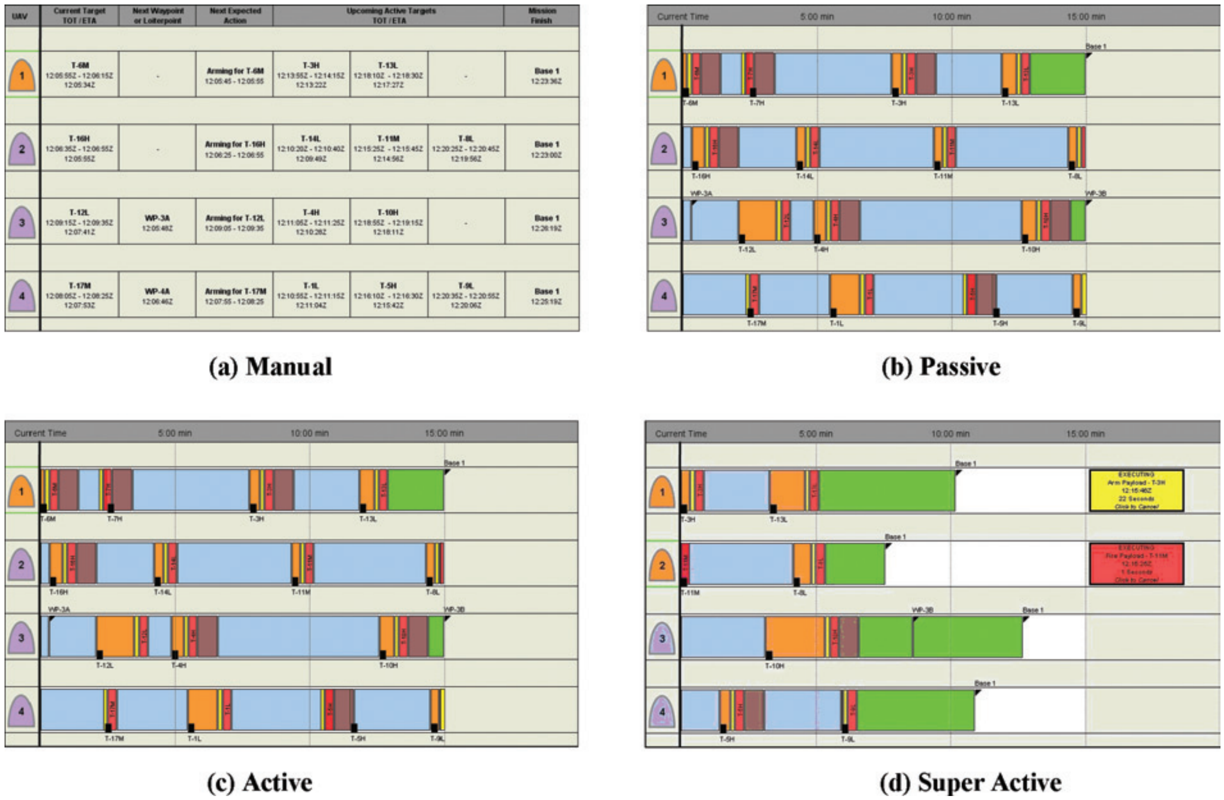
The right-hand side of the MAUVE simulation in Figure 2 provides decision support, and consists of a UAV status window, chat box, UAV health and status updates, and the decision support window. The status window at the top left of the decision support display gives operators low level, detailed information 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,[12] and is an embedded secondary workload and situation awareness measurement tool.[13] 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 a subsequent section. The basic premise of the decision support is to simplify 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 estimated times of arrival (ETAs) for up to the next 4 targets in the 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 tells 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.

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



(a) Manual (b) Passive

(c) Active (d) Super Active

Fig. 3 The four possible levels of decision support in MAUVE.

between the passive and the manual level is graphical schedule integration for users instead of requiring them to piece it together from text, 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

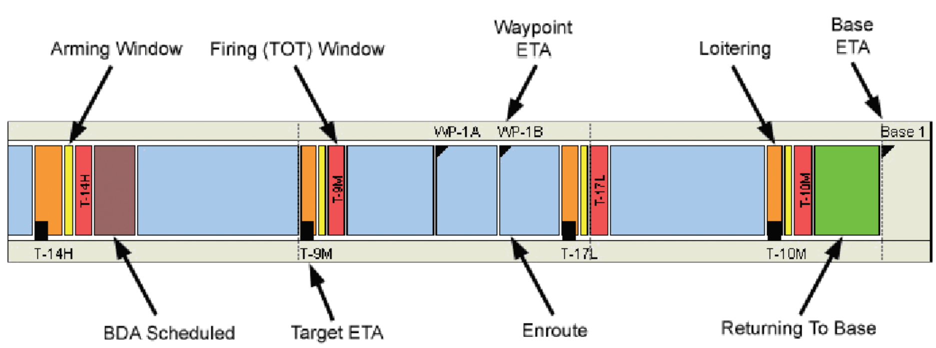


Fig. 4 A representative visual timeline.

instead of specific times to one another. 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[5] demonstrated that even with preview in time-critical tasks, the time participants 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 participants 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 removed that particular schedule conflict, but sometimes created another in the process.

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 is due to fire on, already armed, and the time within the target's pre-defined TOT window. 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 replanning actions and manipulating routes to ensure the UAVs arrive at the correct targets on time. Up to 30 seconds in advance before every arming and firing action, exception boxes 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 and increase in workload would affect operator performance and situational awareness.

A. Apparatus, Participants, and Procedure

Training and testing were conducted on a four screen system called the multi-modal workstation (MMWS),[14] 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 in. and were run at 1280 × 1024 pixels, 16-bit color resolution, while the bottom screen was 15 in. and was run at 1024 × 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. Participants 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 a data file.

A total of 12 participants took part in this experiment, 10 men and 2 women. Subjects were recruited based on whether they had UAV, military and/or pilot experience. The subject population consisted of a combination of students, both undergraduates and graduates, as well as those from the local reserve officer training corps (ROTC) and active duty military personnel. All were paid $10 an hour for their participation. In addition, a $50 incentive prize was offered for the best performer in the experiment. The age range of participants was 20–42 years with an average age of 26.3 years. Nine participants were members of the ROTC or active duty USAF officers, including seven 2nd Lieutenants, a Major and a Lieutenant Colonel. While no participants 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 of the current ATO, which changed 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 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 participants' 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.

B. Experimental Design

Two independent variables were of interest in this experiment: level of decision support (Figure 3) and level of replanning. For the replanning level, operators were required to handle unplanned contingencies such as emergent threat areas and targets, new tasking from superiors, which required them 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 replanning were investigated. The low replanning condition contained 7 replanning events, while the high replanning condition contained 13. Groups of replanning events were interspersed at approximately 3 minute intervals, but under the low replanning level these groups only consisted of a single event. Under high replanning, some groups were composed of 2 or 3 events occurring within 60 seconds of each other. The level of decision support was a between-participants variable and the level of replanning was a within-participants repeated variable, so participants were randomly assigned to a LOA factor level but experienced both replanning conditions.

The dependent variables were a performance score, subjective workload, secondary workload, and situation awareness scores. The performance score measured how well participants met the numerous objectives for a test session and 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, while a perfect score of 1000 meant that all mission objectives were met.

Fig. 5 The MAUVE test bed.

Workload was measured through two metrics: secondary task workload, which measures spare mental capacity, and subjective workload. Secondary task workload was measured through response times to secondary tasks introduced though the instant messaging "chat" interface, which has been shown to be an effective technique for measuring workload in command and control settings.[13] To measure subjective workload, after the completion of each test session, participants filled out a modified NASA Task Load Index (TLX) subjective workload rating survey. The NASA TLX rating scale has been tested in numerous experimental conditions and has been found to be a reliable indicator of subjective workload.[15] The standard NASA TLX procedure computes a single workload score from participants' weighted ratings on a 1–20 scale along six dimensions, which are mental demand, physical demand, temporal demand, effort, performance, and frustration. As this experimental task required no physical demand, subjective workload scores were based upon the five remaining dimensions.

Situation awareness was captured through a subjective SA scale constructed from expert observer ratings. 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.[16] 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.[16,17] Subjective SA scales based upon expert observer ratings have been found to be a reliable and valid measure way to measure SA.[18] This expert rating scale was based upon threat area incursions, system wait time at targets, number of targets missed, and percentage of replanning events successfully completed. Threat area incursions and system wait times measured level 2 SA (or a lack thereof), while missed targets and successful replanning events constituted level 3 SA measures.

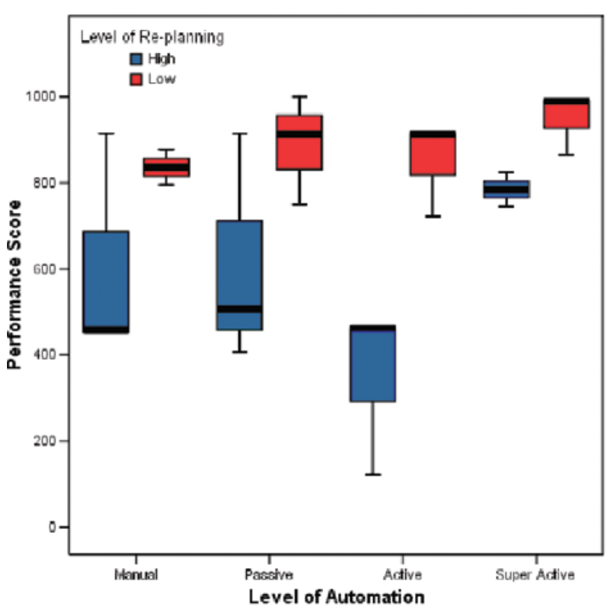
V. Results

For the statistical analysis of the performance scores, a $2 \times 4(3)$ repeated measures linear mixed model was used. Three participants were nested within each of the four automation levels, and the two levels of replanning (within-participants) 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 replanning 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.

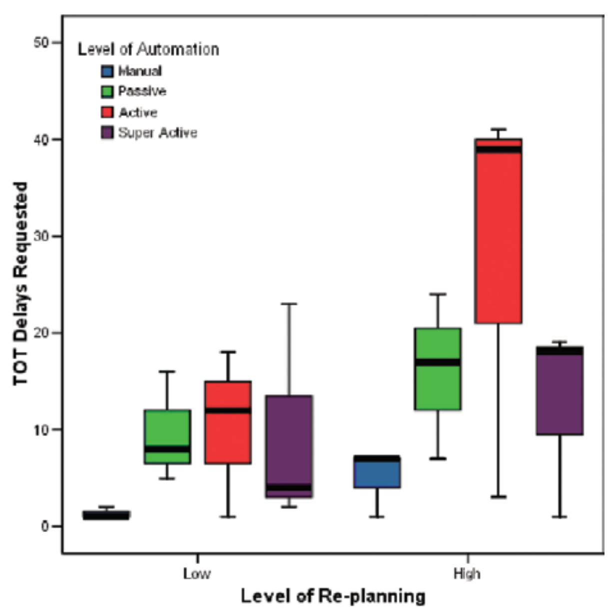
Based on the results from the performance score analysis, more investigation was needed to determine why the active level of automation produced such poor scores. The significant driver of this problem was found to be operators' use of the TOT delay function. Through requesting TOT delays, operators 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 6 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 replanning (Wilcoxon Rank Sum Test, $p = 0.015$), and a higher number of TOT requests in the active level, predominantly for high replanning ($p = 0.065$).

Secondary workload was significantly higher (and thus spare mental capacity significantly less) for the high level of replanning ($F(1, 11.5) = 14.69$, $p = 0.003$), which was expected since workload was essentially doubled. There were significant differences across the levels of automation ($F(3, 12.6) = 9.22$, $p = 0.002$), with significantly higher secondary workload under the manual condition across both levels of replanning. There were no significant interactions between levels of automation and replanning levels. Similarly, subjective workload as measured by the NASA TLX was significantly different across replanning conditions ($F(1, 14.9) = 24.16$, $p < 0.001$), however it was not significant across automation levels. Despite the lack of statistical significance, the highest subjective workload occurred under the active automation condition.

A major consideration in the evaluation of all HSC systems is situation awareness. For level 2 SA measures indicating how well participants comprehended the current situation, a Wilcoxon Signed Rank Test showed a marginally significant difference between levels of replanning ($p = 0.065$) 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, although statisitically, this was a marginal finding. As depicted in Figure 7, participants with the active level



(a)



(b)

Fig. 6 Performance score & TOT request results.

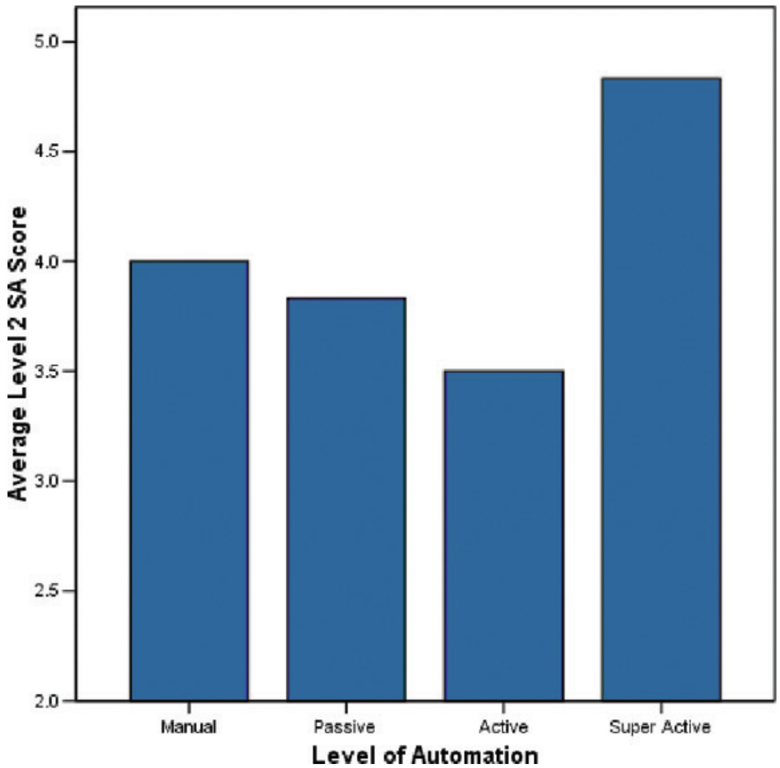


Fig. 7 Level 2 SA results.

of automation had the lowest level 2 SA. For level 3 SA measures that captured participants' ability to project the current situation into the future, a Wilcoxon Signed Rank test demonstrated that participants' level 3 SA for the high replanning scenario was significantly lower ($p = 0.004$) than for the low workload condition, but there was no effect of level of automation.

VI. Discussion

A. Workload and Operator Robustness

The significant difference in performance scores and workload metrics across replanning levels in this experiment indicates that the number of replanning events is an important influence on command and control mission difficulty, as expected. Under low levels of replanning, participants at all levels of automation performed equally well, but under high replanning, participants using active automation had lower performance scores and higher subjective workload scores than all other automation levels. This is an important finding because it demonstrates that operators are fairly robust under low workload regardless of the quality and autonomy of the decision support. However, as will be discussed more in detail in the next section, under high workload operators are much more dependent on external decision aids, and simply adding more visualizations and automated recommendations may make the problem worse.

B. Levels of Automation

The poor overall performance of the active level of automation, indicated by both poor performance scores and poor SA of the participants, was unexpected. It was hypothesized that this level of automated decision support would allow participants to plan ahead more efficiently as demonstrated by Ruff et al.[6] 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. While many previous studies from other domains have also found collaborative types of automation to have superior performance to highly automated or manual systems,[4,19,20] our results showed that management-by-consent caused the worst performance. In this study manual, passive and super active level performance scores were all statistically the same thus no one level outperformed any other level, save for the active which was significantly worse than the other three. A primary limitation in complex systems as demonstrated in this study is that users operating under a management-by-consent design 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 participants' inappropriate use of the "Request TOT Delay" function in the MAUVE simulation. The ability to move a target's TOT was an intervention meant to be used sparingly. Our results demonstrate that under low workload conditions, operators were able to use the "Request TOT Delay" functionality correctly but when faced with high levels of workload, the consequences of requesting a delay were not adequately considered. It is important to highlight that all participants with all levels of automation had the ability to request a TOT change. However, only the participants 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 Delay" function?

C. Global Optimization and Stopping Rule Generation

In the active level of automation, participants clearly were unable to generate appropriate stopping rules when trying to achieve a particular schedule change. 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,[21] it is particularly difficult for them to do under dynamic, non-stationary processes typical of command and control domains.[22,23] Moray et al.[24] found that even if participants 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, participants could not effectively decide when to stop attempting to optimize (minimize) one or more future workload bottlenecks, even though they were warned not to overuse the TOT change request. At the detriment of other tasks and vehicles requiring their attention, participants often focused on obtaining a particular delay until they obtained it. In seeking to globally optimize their schedules, participants were actually narrowly focused on the timeline display, but on the future, not the present. Because this global optimization task was very distracting, 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, participants 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, low probability areas of high workload 15 minutes into the future were generally treated the same as high probability workload areas 1 minute away. Moreover, when participants 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, participants 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 participants 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 TOT 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.[25] Moreover, as will be discussed in the next section, inadequate visualization of the uncertainty could have also been a cause.

D. Preview Visualization

In this study, the problems with stopping rule generation and probabilistic inference were likely 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 and uncertainty. Preview has been shown to lead to erroneous heuristics,[26] while another study found that aiding pilots in attending to lower priority secondary tasks through the use of workload preview had no impact on performance.[27] However, task scheduling results from a process control study suggest the lack of effective preview could be from difficulties in display interpretation,[28] which could also be a factor in this present study. The saliency of the simple reverse-shading technique, while incorporated to allow operators the ability to see the overall timeline in conjunction with a potential high workload area, could have been too high, especially in combination with the absence of more information about the uncertainty. Better representation of uncertainty was not only needed for future workload predictions but also for the likelihood that the TOT change would be approved. Current research with this work is investigating these potential problems and possible design strategies, both graphically and algorithm-based, to improve performance and mitigate bias.

VII. Conclusion

In the quest to design systems such that a single operator can control multiple UAVs, automated decision support tools are a requirement but their impact on operator performance and situation awareness cannot always be predicted. This study, which investigated how levels of automation affect an operator's schedule management strategies when attempting to control multiple independent homogeneous UAVs provides evidence that introducing intelligent aiding can often produce unexpectedly poor results. Moreover the study extends previous work on human supervisory control operator scheduling strategies[5,24] by demonstrating how automated decision aiding and related visualizations can dramatically change these strategies.

The goal of this study was to determine how increasing levels of automation affected operator performance in order to identify possible future automation strategies for multiple UAV scheduling and management. This human-in-the-loop experiment revealed that when operators experienced low workload, increasing automated decision support provided no additional benefit or negative consequence. However, under high workload conditions, the level of automated decision support for schedule management did matter, particularly when an intelligent agent's predictions were included in a visualization. When provided with a high workload preview visualization as well as automated recommendations for workload mitigation, operators attempted to globally optimize their schedules, and did not adequately weigh uncertainty in their decisions. This fixation on the future, which could have been affected by visualization salience, prevented participants 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 intelligent aiding.

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. The goal for computer-aided decision support tool design should be to develop a tool 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, how to better represent uncertainty, and how more sophisticated scheduling algorithms can improve both human and system performance. While human supervisory control decision support tools are typically platform, task, and context-sensitive, identifying prediction visualization techniques and better automated recommendation schemes for time-sensitive operations will benefit not only futuristic multiple UAV operations but also the entire concept of networked command and control operations.

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. and Verplank, W., "Human and Computer Control of Undersea Teleoperators", Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT, Cambridge, MA, 1978.

[2]Chen, T. L. and Pritchett, A. R., Development and Evaluation of a Cockpit Decision-Aid for Emergency Trajectory Generation, *Journal of Aircraft*, 38, 2001, 935–943.

[3]Johnson, K., Ren, L., Kuchar, J., and Oman, C., Interaction of Automation and Time Pressure in a Route Replanning Task, Series "Interaction of Automation and Time Pressure in a Route Replanning Task" *International Conference on Human-Computer Interaction in Aeronautics*, Cambridge, MA, 2002. 132–137.

[4]Layton, C., Smith, P. J., and McCoy, E., Design of a Cooperative Problem-Solving System for En-Route Flight Planning: An Empirical Evaluation, *Human Factors*, 36, 1994, 94–119.

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

[6]Ruff, H., Narayanan, S., and 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.

[7]Wickens, C. D., Dixon, S., and 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.

[8]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.

[9]Kaber, D. B., Endsley, M. R., and 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.

[10]Parasuraman, R., Designing Automation for Human Use: Empirical Studies and Quantitative Models, *Ergonomics*, 43, 2000, 931–951.

[11]Wickens, C. D. and 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.

[12]Cummings, M. L., The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations, *Cyberpsychology and Behavior*, 7, 2004.

[13]Cummings, M. L. and Guerlain, S., Using a Chat Interface as an Embedded Secondary Tasking Tool, Series "Using a Chat Interface as an Embedded Secondary Tasking Tool" *2nd Annual Human Performance, Situation Awareness, and Automation Conference*, Daytona Beach, 2004.

[14]Osga, G. A., Building Goal-Explicit Work Interface Systems, Series "Building Goal-Explicit Work Interface Systems" *ASNE Human Systems Integration Symposium*, Washington DC, 2003.

[15]Wharton, C., Rieman, J., Lewis, C., and Polson, P., The Cognitive Walkthrough Method: A Practitioner's Guide, in Usability Inspection Methods, J. Nielsen and R. Mack Eds., John Wiley & Sons New York, 1994

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

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

[18]Bell, H. H. and Waag, W. L., Using Observer Ratings to Assess Situational Awareness in Tactical Air Environments in Experimental Analysis and Measurement of Situation Awareness, D. J. Garland and M. R. Endsley Eds., 1995.

[19]Vicente, K. J., Work Domain Analysis and Task Analysis, in Cognitive Task Analysis, J. M. Schraagen, S. F. Chipman, and V. L. Shalin Eds., Lawrence Erlbaum Associates Mahwah, NJ, 2000.

[20]Guerlain, S. A., Interactive Advisory Systems, Series "Interactive Advisory Systems" *Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium*, Savannah, Georgia, 2000.

[21]Todd, P. M. and Gigerenzer, G., Precis of Simple Heuristics that Make us Smart, *Behavioral and Brain Sciences*, 23, 2000, 727–780.

[22]Goonetilleke, R. S. and Drury, C. G., Human Optimization with Moving Optima, *Ergonomics*, 32, 1989, 1207–1226.

[23]Klein, G., Analysis of Situation Awareness from Critical Incident Reports, in Situation Awareness Analysis and Measurement, M. Endsley and D. J. Garland Eds., Lawrence Erlbaum Associates Mahwah, NJ, 2000, 51–71.

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

[25]Tversky, A. and Kahneman, D., Judgment under Uncertainty: Heuristics and Biases, *Science*, 185, 1974, 1124–1131.

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

[27]Segal, L. D. and Wickens, C., TASKILLAN II: Pilot Strategies for Workload Management, Series "TASKILLAN II: Pilot Strategies for Workload Management" *Human Factors Society 34th Annual Meeting*, 1990.

[28]Sanderson, P., The Human Planning and Scheduling Role in Advanced Manufacturing Systems: An Emerging Human Factors Domain, *Human Factors*, 31, 1989, 635–666.

James Neidhoefer
Associate Editor