

Predicting Operator Capacity for Supervisory Control of Multiple UAVs

M.L. Cummings, C. E. Nehme, J. Crandall

Humans and Automation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts

With reduced radar signatures, increased endurance and the removal of humans from immediate threat, uninhabited (also known as unmanned) aerial vehicles (UAVs) have become indispensable assets to militarized forces. UAVs require human guidance to varying degrees and often through several operators. However, with current military focus on streamlining operations, increasing automation, and reducing manning, there has been an increasing effort to design systems such that the current many-to-one ratio of operators to vehicles can be inverted. An increasing body of literature has examined the effectiveness of a single operator controlling multiple uninhabited aerial vehicles. While there have been numerous experimental studies that have examined contextually how many UAVs a single operator could control, there is a distinct gap in developing predictive models for operator capacity. In this chapter, we will discuss previous experimental research for multiple UAV control, as well as previous attempts to develop predictive models for operator capacity based on temporal measures. We extend this previous research by explicitly considering a cost-performance model that relates operator performance to mission costs and complexity. We conclude with a meta-analysis of the temporal methods outlined and provide recommendation for future applications.

Introduction

With reduced radar signatures, increased endurance and the removal of humans from immediate threat, uninhabited (also known as unmanned) ae-

rial vehicles (UAVs) have become indispensable assets to militarized forces around the world, as proven by the extensive use of the Shadow and the Predator in recent conflicts.

Current UAVs require human guidance to varying degrees and often through several operators. For example, the Predator requires a crew of three to be fully operational. However, with current military focus on streamlining operations and reducing manning, there has been an increasing effort to design systems such that the current many-to-one ratio of operators to vehicles can be inverted (e.g., [1]). An increasing body of literature has examined the effectiveness of a single operator controlling multiple UAVs. However, most studies have investigated this issue from an experimental standpoint, and thus they generally lack any predictive capability beyond the limited conditions and specific interfaces used in the experiments.

In order to address this gap, this chapter first analyzes past literature to examine potential trends in supervisory control research of multiple uninhabited aerial vehicles (MUAVs). Specific attention is paid to automation strategies for operator decision-making and action. After the experimental research is reviewed for important “lessons learned”, an extension of a ground unmanned vehicle operator capacity model will be presented that provides predictive capability, first at a very general level and then at a more detailed cost-benefit analysis level. While experimental models are important to understand what variables are important to consider in MUAV control from the human perspective, the use of predictive models that leverage the results from these experiments is critical for understanding what system architectures are possible in the future. Moreover, as will be illustrated, predictive models that clearly link operator capacity to system effectiveness in terms of a cost-benefit analysis will also demonstrate where design changes could be made to have the greatest impact.

Previous Experimental Multiple UAV studies

Operating a US Army Hunter or Shadow UAV currently requires the full attention of two operators: an AVO (Aerial Vehicle Operator) and a MPO (Mission Payload Operator), who are in charge respectively of the navigation of the UAV, and of its strategic control (searching for targets and monitoring the system). Current research is aimed at finding ways to reduce workload and merge both operator functions, so that only one operator is required to manage one UAV. One solution investigated by Dixon et al. consisted of adding auditory and automation aids to support the potential single operator [2]. Experimentally, they showed that a single operator

could theoretically fully control a single UAV (both navigation and payload) if appropriate automated offloading strategies were provided. For example, aural alerts improved performance in the tasks related to the alerts, but not others. Conversely, it was also shown that adding automation benefited both tasks related to automation (e.g. navigation, path planning, or target recognition) as well as non-related tasks. However, their results demonstrate that human operators may be limited in their ability to control multiple vehicles which need navigation and payload assistance, especially with unreliable automation. These results are concordant with the single-channel theory, stating that humans alone cannot perform high speed tasks concurrently [3, 4]. However, Dixon et al. propose that reliable automation could allow a single operator to fully control two UAVs.

Reliability and the related component of trust is a significant issue in the control of multiple uninhabited vehicles. In another experiment, Ruff et al. [5] found that if system reliability decreased in the control of multiple UAVs, trust declined with increasing numbers of vehicles but improved when the human was actively involved in planning and executing decisions. These results are similar to those experimentally found by Dixon et al. in that systems that cause distrust reduce operator capacity [6]. Moreover, cultural components of trust cannot be ignored. Tactical pilots have expressed inherent distrust of UAVs as wingmen, and in general do not want UAVs operating near friendly forces [7].

Reliability of the automation is only one of many variables that will determine operator capacity in MUAV control. The level of control and the context of the operator's tasks are also critical factors in determining operator capacity. Control of multiple UAVs as wingmen assigned to a single seat fighter has been found to be "unfeasible" when the operator's task was primarily navigating the UAVs and identifying targets [8]. In this experimental study, the level of autonomy of the vehicles was judged insufficient to allow the operator to handle the team of UAVs. When UAVs were given more automatic functions such as target recognition and path planning, overall workload was reduced.

In contrast to the previous UAVs-as-wingmen experimental study [6] that determined that high levels of autonomy promotes overall performance, Ruff et al. [5] experimentally determined that higher levels of automation can actually degrade performance when operators attempted to control up to four UAVs. Results showed that management-by-consent (in which a human must approve an automated solution before execution) was superior to management-by-exception (where the automation gives the operator a period of time to reject the solution). In their scenarios, their implementation of management-by-consent provided the best situation

awareness ratings and the best performance scores for controlling up to four UAVs.

These previous studies experimentally examined a small subset of UAVs and beyond showing how an increasing number of vehicles impacted operator performance, they were not attempting to predict any maximum capacity. In terms of actually predicting how many UAVs a single operator control, there is very little research. Cummings and Guerlain [9] showed that operators could experimentally control up to 12 Tactical Tomahawk missiles given significant missile autonomy. However, these predictions are experimentally-based which limits their generalizability. Given the rapid acquisition of UAVs in the military, which will soon follow in the commercial section, predictive modeling for operator capacity will be critical for determining an overall system architecture. Moreover, given the range of vehicles with an even larger subset of functionalities, it is critical to develop a more generalizable predictive modeling methodology that is not solely based on expensive human-in-the-loop experiments, which are particularly limited for application to revolutionary systems.

In an attempt to address this gap, in the next section of this paper, we will extend a predictive model for operator capacity in the control of unmanned ground vehicles to a UAV domain [10], such that it could be used to predict operator capacity, regardless of vehicle dynamics, communication latency, decision support, and display designs.

Predicting Operator Capacity through Temporal Constraints

While little research has been published concerning the development of a predictive operator capacity model for UAVs, there has been some previous work in the unmanned ground vehicle (robot) domain [10, 11]. Coining the term “fan-out” to mean the number of robots a human can effectively control, Olsen et al. [9, 22] propose that the number of homogeneous robots or vehicles a single individual can control is given by:

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

In this equation, FO (fan-out) is dependent on NT (Neglect Time), the expected amount of time that a robot can be ignored before its performance drops below some acceptable threshold, and IT (Interaction Time) which is the average time it takes for a human to interact with the robot to ensure it is still working towards mission accomplishment. Figure 1 demonstrates the relationship of IT and NT.

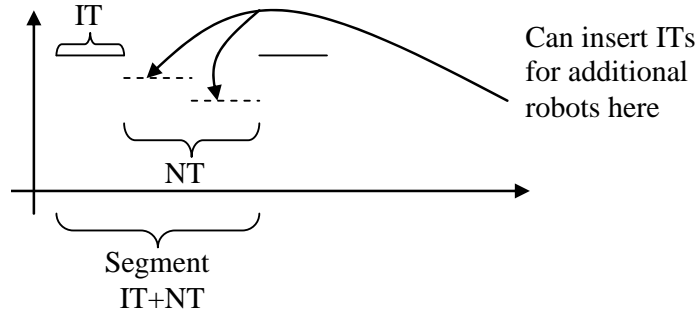


Figure 1: The relationship of NT and IT for a Single Vehicle

While originally intended for ground-based robots, this work has direct relevance to more general human supervisory control (HSC) tasks where operators are attempting to simultaneously manage multiple entities, such as in the case of UAVs. Because the fan-out adheres to Occam’s Razor, it provides a generalizable methodology that could be used regardless of the domain, the human-computer interface, and even communication latency problems. However, as appealing as it is due to its simplicity, in terms of human-automation interaction, the fan-out approach lacks two critical considerations: 1) The important of including wait times caused by human-vehicle interaction, and 2) How to link fan-out to measurable “effective” performance. These issues will be discussed in the subsequent section.

Wait Times

Modeling interaction and neglect times are critical for understanding human workload in terms of overall management capacity. However, there remains an additional critical variable that must be considered when modeling human control of multiple robots, regardless of whether they are on the ground or in the air, and that is the concept of Wait Time (WT). In HSC tasks, humans are serial processors in that they can only solve a single complex task at a time [3, 4], and while they can rapidly switch between cognitive tasks, any sequence of tasks requiring complex cognition will form a queue and consequently wait times will build. Wait time occurs when a vehicle is operating in a degraded state and requires human intervention in order to achieve an acceptable level of performance. In the context of a system of multiple vehicles or robots, wait times are significant in that as they increase, the actual number of vehicles that can be effectively controlled decreases, with potential negative consequences on overall mission success.

Equation 2 provides a formal definition of wait time. It categorizes total system wait time as the sum of the interaction wait times, which are the portions of IT that occur while a vehicle is operating in a degraded state (WTI), wait times that result from queues due to near-simultaneous arrival of problems (WTQ), plus wait times due to operator loss of situation awareness (WTSA). An example of WTI is the time that an unmanned ground vehicle (UGV) idly waits while a human replans a new route. WTQ occurs when a second UGV sits idle, and WTSA accumulates when the operator doesn't even realize a UGV is waiting. In (2), X equals the number of times an operator interacts with a vehicle while the vehicle is in a degraded state, Y indicates the number of interaction queues that build, and Z indicates the number of time periods in which a loss of SA causes a wait time. Figure 2 further illustrates the relationship of wait times to interaction and neglect times.

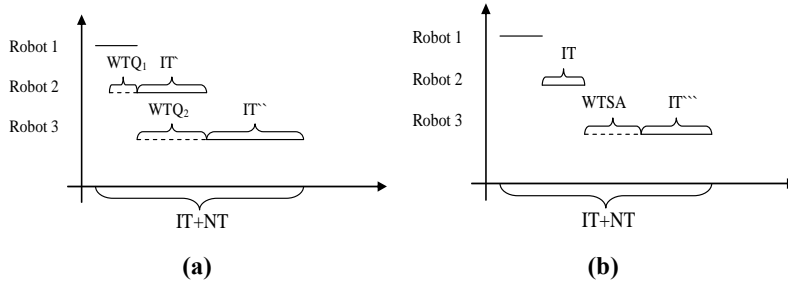


Figure 2: Queuing wait times (a) versus situational awareness wait times (b)

Increased wait times, as defined above, will reduce operator capacity, and Equation 3 demonstrates one possible way to capture this relationship. Since WTI is a subset of IT, it is not explicitly included (although the measurement technique of IT will determine whether or not WTI should be included in the denominator.)

$$WT = \sum_{i=1}^X WTI_i + \sum_{j=1}^Y WTQ_j + \sum_{k=1}^Z WTSA_k \quad (2)$$

$$FO = \frac{NT}{IT + \sum_{j=1}^Y WTQ_j + \sum_{k=1}^Z WTSA_k} + 1 \quad (3)$$

While the revised fan-out (3) includes more variables than the original version, the issue could be raised that the additional elements may not provide any meaningful or measurable improvement over the original equa-

tion which is simpler and easier to model. Thus to determine how this modification affects the fan-out estimate, we conducted an experiment with a UAV simulation test bed, holding constant the number of vehicles a person controlled. We then measured all times associated with equations 1 and 3 to demonstrate the predictions made by each equation. The next section will describe the experiment and results from this effort.

Experimental Analysis of the Fan-out Equations

In order to study operator control of multiple UAVs, a dual screen simulation test bed named the Multi-Aerial Unmanned Vehicle Experiment (MAUVE) interface was developed (Fig. 3). This interface allows an operator to effectively supervise four independent homogeneous 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. The left side of the display provides geo-spatial information as well as a command panel to redirect individual UAVs. The right side of the display provides temporal scheduling decision support in addition to data link “chat windows” commonly in use in the military today [12]. Details of the display design such as color mappings and icon design are discussed elsewhere [13].

The four UAVs launched with a pre-determined mission plan, so initial target assignments and routes were already completed. The operator’s pri-

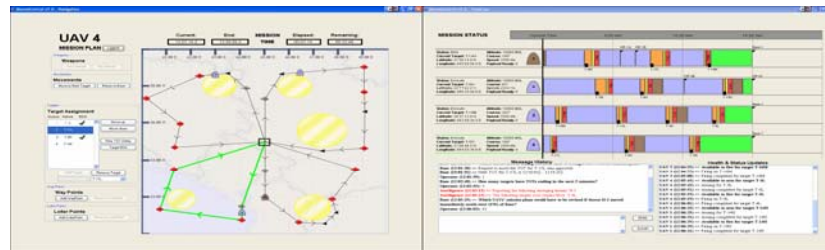


Figure 3: The MAUVE Dual Screen Interface

mary job in the MAUVE simulation was to monitor each UAV’s progress, replan aspects of the mission in reaction to unexpected events and in some cases manually execute mission critical actions such as arming and firing of payloads. The UAVs supervised by participants in MAUVE were capable of 6 high-level actions: traveling enroute to targets, loitering at specific locations, arming payloads, firing payloads, performing battle damage assessment, and returning to base, generally in this order.

In the MAUVE simulations, flight control was fully automated as was the basic navigation control loop in terms of heading control. Operators were occasionally required to replan route segments due to pop-up threat areas so the navigation loop was only partially automated. As will be discussed in more detail next, the mission management autonomy was varied as an independent fact in the experiment.

Levels of Autonomy

Recognizing that the level of autonomy introduced in the mission/payload management control loop can significantly impact an operator's ability to control multiple vehicles, and thus neglect, interaction, and wait times, we developed four increasing levels of decision support for the temporal management of the four UAVs: Manual, Passive, Active, and Super-active, which loosely correlate to the Sheridan and Verplank Levels [14] of 1, 2, 4, 6 (shown in Table 1). The manual level of decision support (Fig. 4a) presents all required mission planning information in a text-based table format. It essentially provides tabular data such as waypoints, expected time on targets, etc., with no automated decision support. It is representative of air tasking orders that are in use by military personnel today.

The passive LOA (Fig. 4b) represents an intermediate mission management LOA in that it provides operators with a color-coded timeline for the expected mission assignments 15 minutes in the future. With this visual representation, recognizing vehicle states with regard to the current schedule is perceptually-based, allowing users to visually compare the relative location of display elements instead of requiring individual parameter searches such as what occurs in the manual condition.

The active LOA (Fig. 4c) uses the same horizontal timeline format as the passive automation level, but provides intelligent aiding. 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. High workload areas, or "bottlenecks," are highlighted through a reverse shading technique while the rest of the colors are muted, but still visible. 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 Time on Target (TOT).

The super-active LOA (Fig. 4d) 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 when the rules of engagement for such actions are met, unless vetoed by the operator in less than 30 seconds (LOA 6, Table 1).

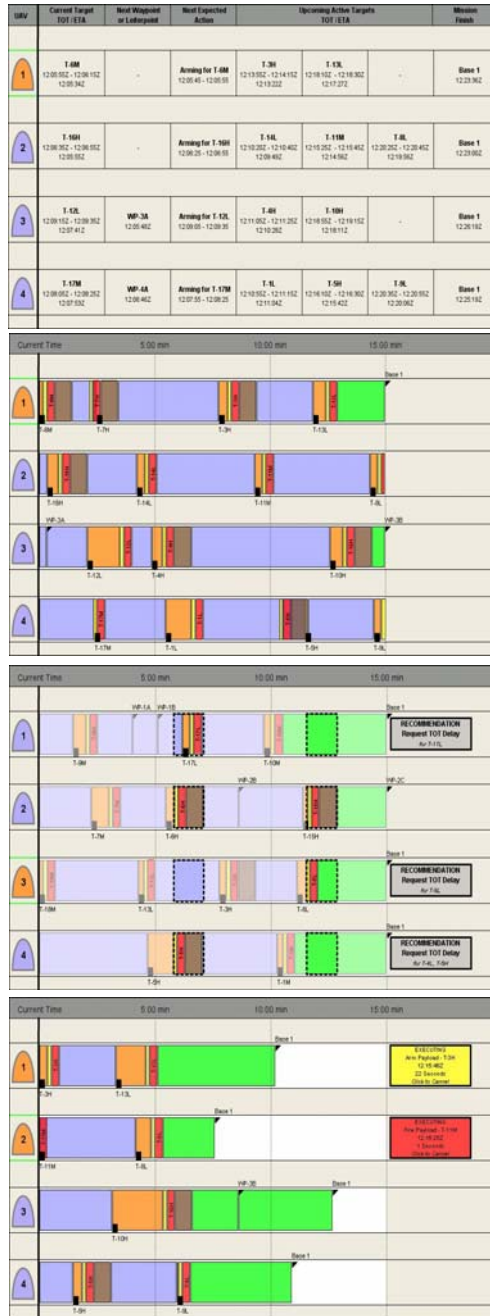
Table 1: Levels of Automation

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.

Experiment Protocol

Training and testing of participants was conducted on a four screen system called the multi-modal workstation (MMWS) [15], originally designed by the Space and Naval Warfare (SPAWAR) Systems Center. The workstation is powered by a Dell Optiplex GX280 with a Pentium 4 processor and an Appian Jeronimo Pro 4-Port graphics card. During testing, all mouse clicks, both in time and location, 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 participants took part in this experiment, 10 men and 2 women, and they were recruited based on whether they had UAV, military and/or pilot experience. The participant 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/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.

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, and responding to online instant messages. Following training, participants tested on two consecutive 30 minute sessions, which represented low and high workload scenarios. These were randomized and counter-balanced to prevent a possible learning effect. The low replanning condition contained 7 replanning events, while the high replanning condition contained 13. Each simulation was run several times faster than real time so an entire strike could take place over 30 minutes (instead of several hours).

Figure 4 a- d: Increasing Levels of Decision and Action Support in Mauve

Results and Discussion

In order to determine whether or not the revised fan-out prediction in (3) provided a more realistic estimate than the original fan-out (1), the number of vehicles controlled in the experiment was held constant (four) across all levels of automation. Thus if our proposed prediction was accurate, we should be able to predict the actual number of vehicles the operators were controlling. As previously discussed, all times were measured through interactions with the interface which generally included mouse movements, selection of objects such as vehicles and targets for more information, commanding vehicles to change states, and the generation of communication messages.

Neglect time was counted as the time when operators were not needed by any single vehicle, and thus were monitoring the system and engaging in secondary tasks such as responding to communications. Because loiter paths were part of the preplanned missions, oftentimes to provide for buffer periods, loiter times were generally counted as neglect times. Loitering was only counted as a wait time when a vehicle was left in a loiter pattern past a planned event due to operator oversight. Interaction time was counted as any time an operator recognized that a vehicle required intervention and specifically worked towards resolving that task. This was measured by mouse movements, clicks, and message generations. The method of measuring NT and IT, while not exactly the same as [11], was driven by experimental complexity in representing a more realistic environment. However, the same general concepts apply in that neglect time is that time when each vehicle operated independently and interaction time is that time one or more vehicles required operator attention.

As discussed previously, wait times were only calculated when one or more vehicle required attention. Wait time due to interactions (e.g., the time it took an operator to replan a new route once a UAV penetrated a threat area) was subsumed in interaction time. Wait time due to queuing occurred when, for example, a second UAV also required replanning to avoid an emergent threat and the operator had to attend to the first vehicle's problem before immediately moving to the second. Wait time due to the loss of situation awareness was measured when one or more vehicles required attention but was not noticed by the operator. This was the most difficult wait time to capture since operators had to show clear evidence that they did not recognize a UAV required intervention. Examples of wait time due to loss of SA include the time UAVs spend flying into threat areas with no path correction, and leaving UAVs in loiter patterns when they should be redirected.

Figures 5 and 6 demonstrate how the wait times varied both between the two fan-out equations as well the increasing levels of automation under

low and high workload conditions respectively. Using the interaction, neglect, and wait times calculated from the actual experiment, the solid line represents the predictions using (1), the dashed line represents the predictions of (2), and the dotted line shows how many UAVs the operators were actually controlling, which was held constant at four.

Low Workload Predictions

Under the low workload condition, three important trends should be noted. Under the lower levels of automation for both the original and revised fan-out equations, operator capacity was essentially flat, and a significant increase was not seen until the use of a higher automation strategy, management-by-exception, was implemented. It is important to remember that the metric is *time* and not overall decision quality or performance. However, independent performance measures indicated that at the low workload level, operators were able to effectively control all four vehicles [16].

The second trend of note is the fact that for the low workload condition, the revised fan-out model (3) provides a more conservative estimate of approximately 20% under that of the model that does not consider wait times (1). However, the third important trend in this graph demonstrates that for both (1) and (3) the predictions were much higher than the actual number of UAVs controlled. This spare capacity under the low workload condition was empirically observed, in that subjective workload measures (NASA-TLX) and performance scores were statistically the same when compared

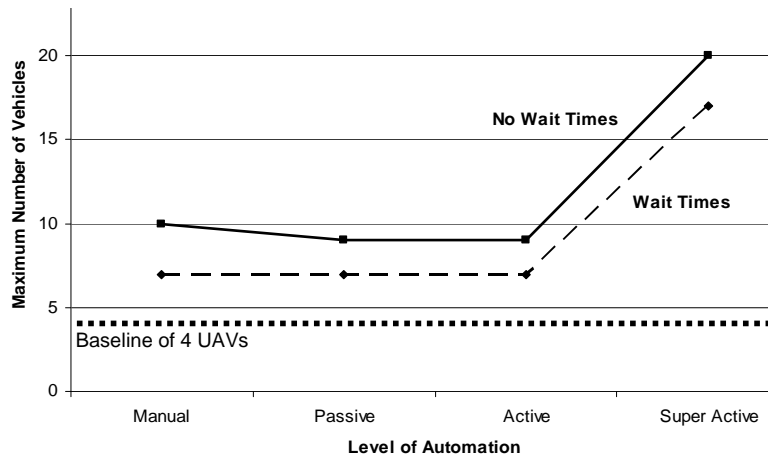


Figure 5: Low Workload Operator Capacity Predictions

across all four levels of autonomy (lowest pair wise comparison p value = .111 ($t = 1.79$, $DOF = 8$), and $p = .494$ ($t = .72$, $DOF = 8$) respectively).

Thus for the low workload condition across all levels of automation, operators were underutilized and performing well. Thus they theoretically could have controlled more vehicles. Using the revised FO model (3) and under the manual, passive, and active condition, operators' theoretical capacity could have increased by ~75% (up to 7 vehicles). Under the highest autonomy for mission management, predictions estimate operators could theoretically control as an upper limit, four times as many, ~17 vehicles [17]. Previous air traffic control (ATC) studies have indicated that 16-17 aircraft are the upper limit for enroute air traffic controllers. Since controllers are only providing navigation assistance and not interacting with flight controls and mission sensors (such as imagery), the agreement between ATC enroute controller capacities and low workload for UAV operators is not surprising.

High Workload Predictions

While the low workload results and predictions suggest that operators are capable of controlling more than four vehicles in MAUVE, the results from the high workload scenarios paint an entirely different picture. The high workload scenarios were approximately double the workload over the low workload scenarios, and represent a worst case scenario. Performance results indicate that those operators with the active level of automation were not able to control their four UAVs effectively, but all other operators were with varying degrees of success. As in the low workload condition, the revised fan-out model (3) is the more conservative and as demonstrated in Figure 6, more closely predicts the actual number of four vehicles assigned to each operator. Moreover, while under the low workload condition, the estimates of controller capacity dropped almost uniformly across automation levels by 20% for the original fan-out model. However, under high

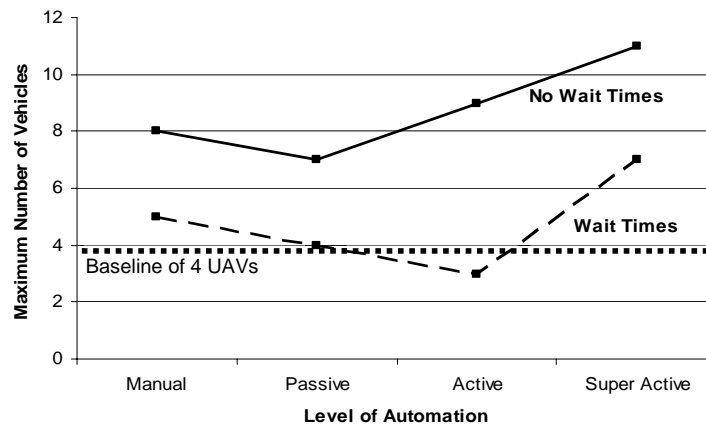


Figure 6: High Workload Operator Capacity Predictions

workload, they dropped 36-67% for the model that includes wait times. The largest difference between conditions occurred for the active level of automation. In addition to the lower number of predicted vehicles, the active condition produced statistically lower performance scores (e.g., $t = 2.26$, $DOF = 8$, $p = 0.054$ for the passive-active comparison). This was attributed to the inability of subjects in the active condition to correctly weight uncertainty parameters and is discussed in detail elsewhere [16].

As in the low workload results, subjects performed the best (in terms of time management) under the highest level of automation for mission management (super-active), with a theoretical maximum of seven vehicles. However, under this condition in the experiment, subjects exhibited automation bias and approved the release of more weapons on incorrect targets than for the passive and active levels. Automation bias, the propensity for operators to take automated recommendations without searching for disconfirming evidence, has been shown to be a significant problem in command and control environments and also operationally for the Patriot missile [18]. Thus increased operator capacity for management-by-exception systems must be weighed against the risk of incorrect decisions, by either the humans or the automation.

Wait Time Proportions

Figures 5 and 6 demonstrate that the inclusion of wait times in a predictive model for operator capacity in the control of MUAVs can radically reduce the theoretical maximum limit. Figure 7 demonstrates the actual proportions of wait time that drove those results. Strikingly, under both low and high workload conditions, the wait times due to the loss of situation awareness dominated overall wait times.

This partitioning of wait time components is important because it demonstrates where and to what degree interventions could potentially improve both human and system performance. In the case of the experiment detailed in this chapter, clearly more design intervention, from both an automation and HCI perspective, is needed that aids operators in recognizing that vehicles need attention. As previously demonstrated, some of the issues are directly tied to workload, i.e., operators who have high workloads have more loss of SA. However, often loss of SA occurred because operators did not recognize a problem which could be mitigated through better decision aiding and visualization.

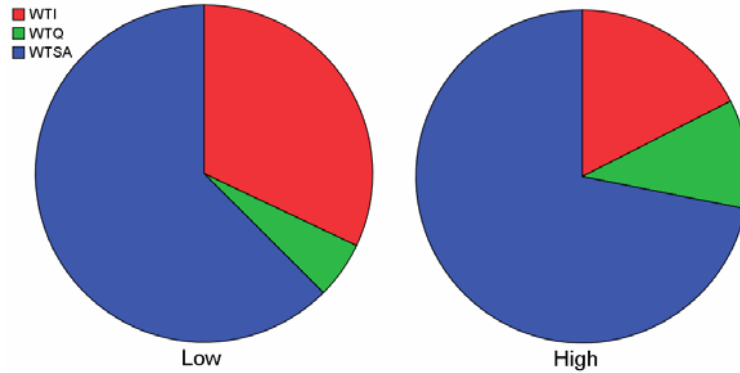


Figure 7: Wait Time Proportions

Linking Fan-out to Operator Performance

Results from the experiment conducted to compare (1) AND (3), the revised fan-out model which includes wait times is both more conservative and closer to the actual number of vehicles under successful control. While under low workload, both the experiment and prediction indication that operators could have controlled more vehicles than four, the only high workload scenario in which operators demonstrated any spare capacity was with the super-active (management-by-exception) decision support. Moreover, wait time caused by the lack of situation awareness dominated overall wait time. In addition, this research demonstrates that both workload and automated decision support can dramatically affect wait times and thus, operator capacity.

While more pessimistic than the original fan-out equation (1), the revised fan-out equation can really only be helpful for broad “ballpark” predictions of operator capacity. This methodology could provide system engineers with a system feasibility metric for early manning estimations, but what primarily limits either version of the fan-out equation is its inability to represent any kind of cost trade space. Theoretically fan-out, revised or otherwise, will predict the maximum number of vehicles an operator can *effectively* control, but what is effective is often a dynamic constraint. Moreover, the current equations for calculating fan-out do not take into account explicit performance constraints. In light of the need to link fan-out to some measure of performance, as well as the inevitability of wait times introduced by human interaction, we propose that instead of a simple

maximum limit prediction, we should instead find the optimal number of UAVs such that the mission performance is maximized.

The Overall Cost Function

Maximizing UAV mission performance is achieved when the overall performance of all of the vehicles, or the team performance, is maximized. Consider multiple UAVs that need to visit multiple targets, either for destruction (SEAD missions as discussed previously) or imaging (typical of Intelligence, Search, and Reconnaissance (ISR) missions). A possible cost function is expressed in (4):

$$C = \text{Total_Fuel_Cost} + \text{Total_Cost_of_Missed_Targets} + \text{Total_Operational_Cost} \quad (4)$$

Total_Fuel_Cost is the amount of fuel spent by all the vehicles for the duration of the mission multiplied by the cost of consuming that fuel. The Total_Cost_of_Missed_Targets is the number of targets not eliminated by any of the UAVs multiplied by the cost of missing a single target. The Total_Operational_Cost is the total operation time for the mission multiplied by some operational cost per time unit, which would include costs such as maintenance and ground station operation costs. This more detailed cost function is given in (5).

$$C = \text{cost_of_fuel} * \text{total_UAV_distance} + \text{cost_per_missed_target} * \#_of_missed_targets + \text{operation_cost_per_time} * \text{total_time} \quad (5)$$

In order to maximize performance, the cost function should be minimized by finding the optimal values for the variables in the cost equation. However, the variables in the cost equation are themselves dependent on the number of UAVs and the specific paths planned for those UAVs. One way to minimize the cost function is to hold the number of UAVs variable constant at some initial value and to vary the mission routes (individual routes for all the UAVs) until a mission plan with minimum cost is found. We then select a new setting for the number of UAVs variable and repeat the process of varying the mission plan in order to minimize the cost. After iterating through all the possible values for the number of UAVs, the number of UAVs with the least cost and the corresponding optimized mission plan are then the settings that minimize the cost equation. As the number of UAVs is increased, new routing will be required to minimize the cost function. Thus, the paths, which determine time of flight, are a function of number of UAVs.

Moreover, if a target is missed, then there is an additional, more significant cost. When the number of UAVs planned is too low, the number of missed targets increases and hence the cost is high. When the number of UAVs is excessive, more UAVs are being used than required and thus additional, unnecessary costs are incurred. We therefore expect the lowest cost to be somewhere in between those two extremities, and that the shape of the cost curve is therefore concave upwards¹ (Figure 8). The profile in Figure 8 does not include the effect of wait times, and it does not take into account the interaction between the vehicles and the human operator.

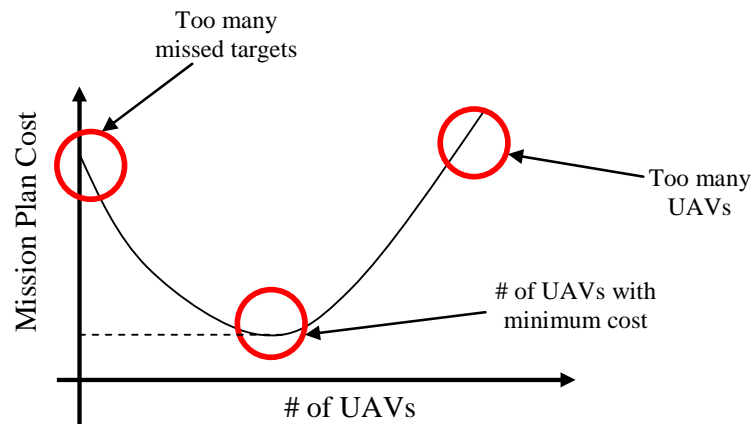


Figure 8: Mission Plan Costs as a Function of Number of UAVs

In terms of wait times, any additional time a vehicle spends in a degraded state will add to the overall cost expressed in (5). Wait times that could increase mission cost can be attributed to 1) Missing a target which could either mean physically not sending a UAV to the required target or sending it outside its established TOT window, and 2) Adding flight time through route mismanagement, which in turn increases fuel and operational costs. Thus, wait times will shift the cost curve upwards. However, because wait times will likely be greater in a system with more events, and hence more UAVs, we expect the curve to shift upwards to a greater extent as the number of UAVs is increased.

In order to account for wait times in a cost-performance model, which as previously demonstrated is critical in obtaining a more accurate operator

¹ Note that this claim is dependent on the assumption that the UAVs independently perform tasks.

capacity prediction, we need a model of the human in our MUAV system, which we detail in the next section.

The Human Model

Since the human operator’s job is essentially to “service” vehicles, one way to model the human operator is through queuing theory. The simplest example of a queuing network is the single-server network shown in Figure 9.

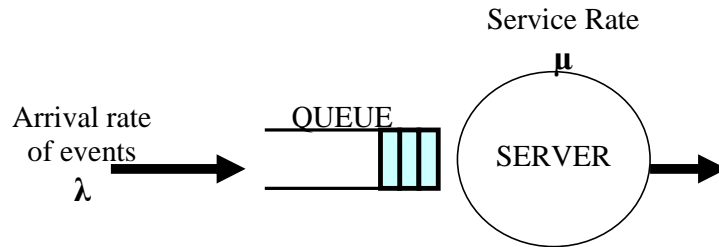


Figure 9: Single Server Queue

Modeling the human as a single server in a queuing network allows us to model the queuing wait times, which can occur when events wait in the queue for service either as a function of a backlog of events or the loss of situation awareness. For our model, we will model the inter-arrival times of the events with an exponential distribution, and thus the arrivals of the events will have a Poisson distribution. In terms of our model, the events that arrive are vehicles that require intervention to bring them above some performance threshold. Thus neglect time for a vehicle is the time between the arrival of events from that particular vehicle and interaction time is the same as the service time.

The arrival rate of events from each vehicle is on average one event per each (NT+IT) segment. The total arrival rate of events to the server (the operator) is the average arrival rate of events from each vehicle multiplied by the number of vehicles.

$$Arrival_rate = \lambda = \#_of_UAVs * \frac{1_event}{(NT + IT)} = \frac{\#_of_UAVs}{NT + IT} \frac{events}{time} \quad (6)$$

In terms of the service rate, by definition, the operator takes, on average, an IT length of time to process each event. Therefore assuming that the operator can constantly service events (i.e., does not take a break while events are in the queue):

$$Service_rate = \mu = \frac{1}{IT} \frac{events}{time} \quad (7)$$

By using Little's theorem, we can show that the mean time an event spends in the queue is:

$$Wq = \frac{\lambda / \mu}{\mu - \lambda} \quad (8)$$

For the purposes of our predictive model, we will assume that this wait time in the queue (Wq , eqn 8) includes both situation awareness wait times (WTSA) as well as wait times due to operator engagement in another task (referred to as WTQ in the previous section).

Now that we have established our operator model based on queuing theory, we will now show how this human model can be used to determine operator capacity predictions through simulated annealing optimization.

Optimization through Simulated Annealing

The model that captures the optimization process for predicting the number of UAVs that a single operator can control is depicted in Figure 10. The optimizer takes in as input the number_of_UAVs, the mission description (including the number of targets and their locations), parameters describing the vehicle attributes (such as UAV speed), and other parameters including the weights that are used to calculate the cost of the mission plan. The optimizer in our model (programmed in MATLAB[®]) iterates through the #_of_UAVs variable, applying a Simulated Annealing algorithm to find the optimal paths plan, as described earlier. The #_of_UAVs with the smallest cost is then selected as that corresponding to the optimal setting. As previously discussed, the human is modeled as a server in a priority queuing system that services events generated by the UAVs according to arrival priorities. The average arrival and service rates as well as their corresponding probabilistic distributions are as assumed earlier.

We chose the simulated annealing (SA) technique for heuristic-based optimization. There were several benefits to selecting the SA technique over other optimization techniques. First, SA is a technique that is well suited to avoiding local minima, a property that is necessary when sub-optimal solutions can exist whilst searching for the global optimum as is the case in evaluating different mission plans. Also, SA introduces randomness such that the technique generates alternative acceptable solutions on different runs, hence allowing the system designer to seek alternative optimal designs when initial solutions are not feasible. Two limitations of

SA are that problems with many constraints can be difficult to implement and that run times can be long. Our problem, however, is one of few constraints and hence their implementation was not an issue. Also, since optimization takes place in mission planning stages and not in time-critical mission replanning, the long run times have a minimal adverse effect.

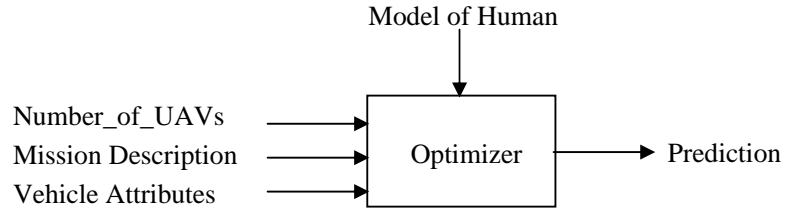


Figure 10: Optimization Model

Model Parameters, Constraints, and Variables

The list of parameters established for the design process is presented in Table 2. We selected generic UAV capabilities that would be exhibited by small-to-medium size UAVs engaged in an ISR mission such as the Hunter or Shadow. Our cost function was discussed previously (5) and Table 3 details the constraints used in our model.

Table 2: Optimization Parameters

Name	Unit	Value
Mission Data (includes number of targets, time on targets, and locations)	-	5-10 targets
UAV speed	mi/hr	100
UAV Endurance	hr	5
UAVs launch location	Cartesian	0,0
Cost_per_missed_target	\$/target	1500
Cost_of_fuel_per_min	\$/min	10
Cost_of_operations_per_min	\$/min	1
NT	min	3 ²
IT	min	0.3 ¹

² Interaction and neglect times were determined using the MAUVE interface described previously.

Table 3: Constraints for Simulated Annealing

Constraints
<ul style="list-style-type: none">• A UAV cannot visit targets for which it cannot meet the times on target• Each UAV must visit at least one target• UAV routes must start and end at launch locations.• The total distance traveled by each UAV cannot exceed a maximum range parameter

Results of Simulation

We first investigated the cost-UAV number relationship for the theoretical best case in which the human operator is “perfect” and introduces no delays in the system. In Figure 11, the optimized cost is plotted against the number of UAVs variable, with a mission plan of 10 targets. As predicted, the curve is concave upwards and has a global minimum where the cost is minimized.

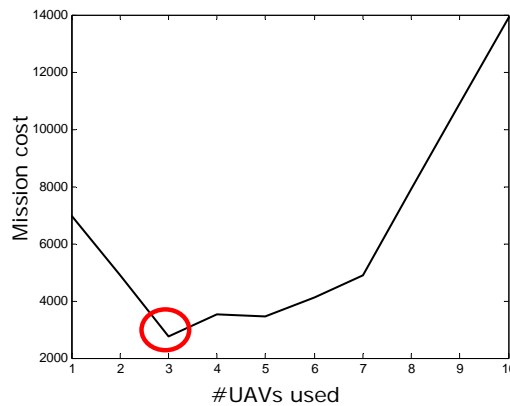


Figure 11: Minimum Cost versus Number of UAVs

We then proceeded to include the effect of the human operator and hence, wait times. It was assumed that during periods of wait times, UAVs loitered in the same spot and therefore maintained the same geographic location on the map. In order to demonstrate how plan complexity could affect the problem, we included three scenarios in which 5, 7, and 10 targets were represented, as depicted in Figure 12.

Figure 12a shows the effect of a relatively simple mission with only 5 targets. In general, the effect of wait times on the cost curve is minimal, i.e., the minimum theoretical best case is equal to that with the operator

wait time case. This is not unexpected, since simple missions have a rather small sensitivity to wait times. The simplicity of the missions does not overburden the operator who is operating in a robust cognitive state, and can accommodate the wait times without incurring increased costs.

Figure 12b demonstrates how the curves can shift as a function of increasing task complexity (7 versus 5 targets). The cost curves between the theoretical best case and the operator wait time model clearly deviate and the optimal number of UAVs that should be controlled decreases from 4 to 3 UAVs. This is primarily due to the fact that the wait times generally affect the longer routes where the probability of missing targets increases.

The results for 10 targets are shown in Figure 12c, which demonstrate a significant divergence from the theoretically perfect operator model and the model with wait times, especially beyond 5 UAVs. Interestingly, as in the case with 5 targets, the minimum cost occurs at 2 UAVs for 10 targets. However, in the case of 5 targets, only 2 UAVs were needed to meet the operational requirements and the operator could meet this demand. How-

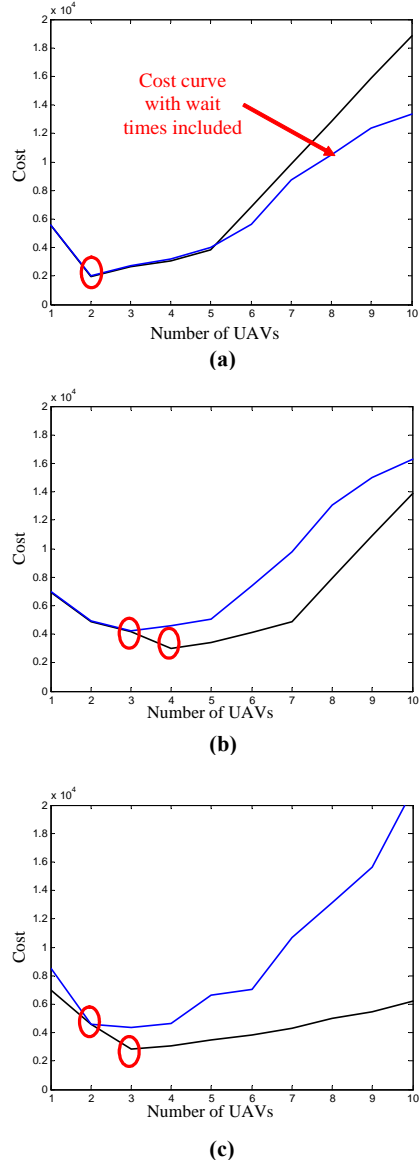


Figure 12: Cost curves with (a) 5 targets (b) 7 targets, and (c) 10 targets

ever, with 10 targets, 2 UAVs became the minimum cost point because of operator limitations, i.e., the wait times incurred by controlling more vehicles became unacceptably high. Thus, at the inflection point in these curves, the left region is primarily constrained by operational demands, but the right region is dominated by human performance limitations, specifically wait times, as seen in Figure 13.

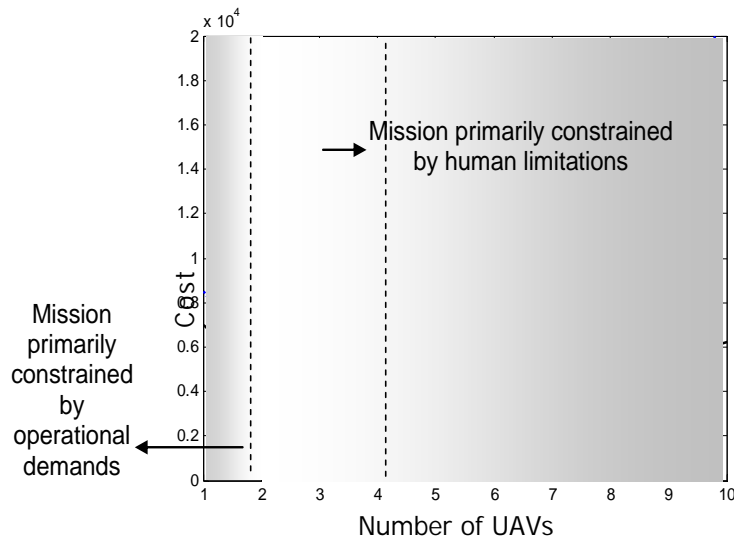


Figure 13: Operational Demands vs. Human Limitations in Mission Planning

Another interesting trend across the predictions in Figures 12a-c is the relatively flatness of the curves in the left region. For example, there is generally a plateau in performance such that dramatic increases are not seen in cost until more than 5 UAVs are managed. In the case of 5 targets and up to 5 UAVs under control, costs increased at a rate of 33% per additional UAV. Beyond 5 UAVs, the mission increased at a rate of 74% per UAV, a much sharper increase, which suggests that operator performance is relatively robust up to 5 UAVs, at which point the operator is saturated and severely limits overall mission success. These graphs demonstrate that the more complex mission requirements added to the cognitive load of the operators, thus workload had to be reduced by reducing the number of UAVs under control.

In terms of the fan-out and revised fan-out equations (1 & 3), using the same neglect and interaction times as in the cost-based simulation model (Table 2), as well as the wait times derived from the queuing theory model

(8), the predictions are seen Figure 14. While the original fan-out estimated would predict a constant 11 vehicles, given the wait times that would build given increasing numbers of UAVs under control, such that the optimal control point is 5 vehicles. Interestingly this number is very close to what was experimentally observed in the previously described experiment.

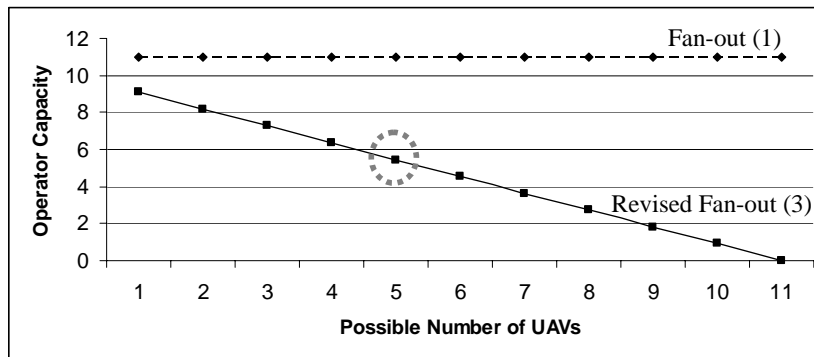


Figure 14: Predictions Using Cost-Based Simulation Inputs

Meta-Analysis of the Experimental and Modeling Prediction methods

Two methods for determining maximum operator capacity for supervisory control of multiple UAVs have been presented, both based on operator interactions and wait times for mission tasks, as well as neglect times during which one or more vehicles operate autonomously. The strengths and weaknesses of each method will now be discussed, as well as how these methods could be used synergistically.

In the first method, the original fan-out equation that related operator interaction and vehicle neglect times (1) was revised to include operator wait times (3). An experiment was conducted to determine if the revised fan-out predictions more closely matched actual human-in-the-loop control scenarios. The results showed that the revised fan-out model produced more conservative estimates when modified to include wait times caused by human interactions, which include interaction wait time, wait time in the queue, and wait time due to the loss of situation awareness.

While this temporal-based method for computing fan-out gives more conservative general estimates, it lacks the cost-benefit analysis trade space representation that can be found through optimization methods that

provide for sensitivity analysis. For example, in the experiment, it was estimated that operators could control 7-16 UAVs in a low workload scenario, but only 3-7 vehicles in high workload settings. The ranges resulted from increasing levels of automation as an experimental independent variable. Because these predictions were based on experimental data (which were discrete across four different levels of automation), there can be no post-hoc sensitivity analysis, only refining the experimental method and running more human subject trials, which is very expensive and labor intensive.

In comparison, optimization methods such as the example presented here provide not only predictions for operator capacity but also directly link the capacity to a system performance measure, which was cost in our example. By developing the estimates through the fan-out approach, there is only the consideration of a vaguely defined threshold for acceptable operator performance. Furthermore, there is no way to directly infer how this human performance affects the overall system, which is actually the more critical variable, particularly in command and control settings. Moreover, while it was very expensive in terms of experimental design for human subjects to examine mission complexity in terms of low and high workload, in the cost-based simulation method, mission complexity was represented by the number of targets, which was relatively not costly to alter. Thus, this type of prediction method allows for more specific and detailed predictions for operator capacity, as well as how the external environment (i.e., number of targets) will affect overall mission success.

However, while the simulation estimations provide for multivariate sensitivity analysis across operator and system performance metrics, one drawback is the inability to directly correlate the predictions to possible design interventions. As previously discussed, the cost-based simulation links the external environment to both operator and system performance, but it inherently lacks the ability to parse out which system parameters could and should be changed to improve operator and autonomy performance. For example, in the SA model, all wait times are included in a single measure, however the wait times (interaction, queuing, and situation awareness) fundamentally have different causes. In addition, as demonstrated in Figure 7, the different types of wait times can have dramatically different values and without the ability to model and see the separate effects of different wait time sources, it is not clear what design interventions could occur to mitigate them (such as improved decision support or increased vehicle autonomy.)

Moreover, a cost-based simulation cannot represent the impact of specific automation strategies on operator performance. It is often assumed that as autonomy levels increase (as depicted in Table 1), the need for human interaction decreases, and thus lowers system wait times. However, as can be seen in Figure 15, these assumptions are not always accurate. In the

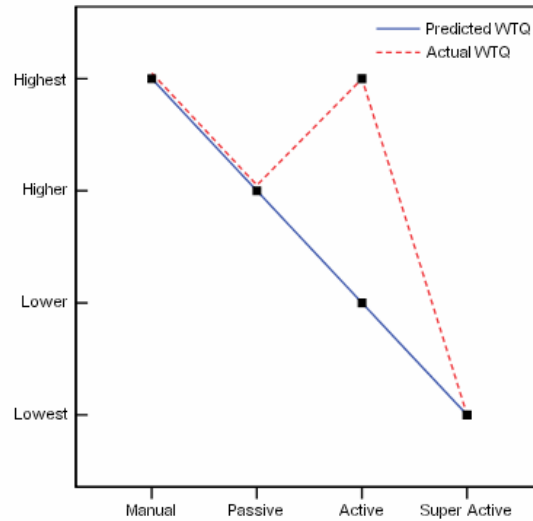


Figure 15: Wait Times in the Queue across Levels of Automation

experiment previously discussed, we predicted that as system autonomy increased, wait times due to an operator workload queue (referred to as wait time in the queue in the previous section) would decrease. However, the dotted line demonstrates what queuing wait times were actually observed, and there was clearly an anomaly with the active condition that corresponds to LOA 4 in Table 1. Described more in detail in [16], what was hypothesized to be a decision support tool to mitigate operator workload actually degraded operator performance and caused increased, instead of decreased, wait times. This insight was only gained through the experimental derived interaction, neglect, and wait times. Because the SA optimization approach and other similar stochastic approaches assume an a priori distribution (both in arrival rates and service times), if such simulation methods are not used in conjunction with experimentally derived data, results are highly speculative and lack external validity.

This last point about the problem with assumptions highlights an inherent limitation to both methods: Estimating interaction, neglect, and wait times. As previously discussed, for the cost-based simulation, a distribution must be selected for wait times, and presently there is little theoretical

or empirical basis for doing so. In addition, interaction and neglect times must be selected a priori and while these could be estimated from system design parameters, they are highly contextual and will likely dramatically change with different levels of autonomy, decision support, mission complexity, operator training, etc.

Similarly, even experimentally derived interaction, neglect, and wait times can be difficult to measure. Unfortunately the times and the associated costs (degraded performance, etc.) are very difficult to capture in performance-based simulations such as the one reported in this study. These difficulties have also been identified in the control of multiple ground robots [26]. Through using software that tracked users' cursor movements and activation of control devices, we were able to determine on a gross level when a subject was actively engaged with a particular UAV, but subtle transitions are difficult to capture. The use of psychophysiological measurement devices may be of use in addition to performance-based measures but needs significantly more investigation.

Conclusions

With the recognition that intelligent autonomy could allow a single operator to control multiple vehicles (including air, ground, and water), instead of the converse which is true today, there is increasing interest in predicting the maximum numbers of autonomous vehicles an operator can control. A critical system architecture question is then how many vehicles could one operator control? While there are other methods that could be used to predict this number (e.g., cognitive modeling which suffers from the ability to represent highly complex systems, and simulations and experiments with advanced prototypes, which suffer from exorbitant development costs), we demonstrated, through two different methods, how this number can be estimated by considering the temporal elements of supervisory control of multiple UAVs.

In the first method, we demonstrated that past equations of fan-out omitted important aspects of human interactions with multiple UAVs. We suggest an alternative equation that captures some of these aspects using wait times. However, these temporal approaches to measuring fan-out are limited since these results are not explicitly linked to performance. In comparison, we used cost-based simulation model that links operator performance to both mission costs and complexity; however, it suffers from problematic assumptions and an inability to highlight specific areas for design interventions.

While each method has strengths and weaknesses, they are not mutually exclusive. The two approaches can be synergistic in that temporal data gathered experimentally for initial rough estimates such as fan-out can provide more valid simulation models. Predictions then made through optimization simulations can be further refined through sensitivity analyses and appropriately focused human-in-the-loop experiments. In this way, effects of increasing UAVs and/or system autonomy can be seen on system performance as well as operator performance. In terms of application, this iterative approach to predicting operator capacity would likely provide the most benefit early in the systems engineering conceptual stages when unmanned aerial systems are still in development and uncertainty in system parameters is high.

Acknowledgments

The research was supported by grants from Boeing Phantom Works, and Lincoln Laboratories. We would also like to recognize the experimental contributions of Paul Mitchell.

References

- [1] J. Franke, V. Zaychik, T. Spura, and E. Alves, "Inverting the Operator/Vehicle Ratio: Approaches to Next Generation UAV Command and Control," presented at Association for Unmanned Vehicle Systems International and Flight International, Unmanned Systems North America Baltimore, MD, 2005.
- [2] S. Dixon, C. Wickens, and D. Chang, "Mission Control of Multiple Unmanned Aerial Vehicles: A Workload Analysis," *Human Factors*, in press.
- [3] A. T. Welford, "The psychological refractory period and the timing of high-speed performance - a review and a theory," *British Journal of Psychology*, vol. 43, pp. 2-19, 1952.
- [4] D. E. Broadbent, *Perception and Communication*. Oxford: Pergamon, 1958.
- [5] H. A. Ruff, S. Narayanan, and M. H. Draper, "Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles," *Presence*, vol. 11, pp. 335-351, 2002.
- [6] S. Dixon, C. D. Wickens, and D. Chang, "Unmanned Aerial Vehicle Flight Control: False Alarms Versus Misses," presented at

- Humans Factors and Ergonomics Society 48th Annual Meeting, New Orleans, 2004.
- [7] M. L. Cummings and D. Morales, "UAVs as Tactical Wingmen: Control Methods and Pilots' Perceptions," in *Unmanned Systems*, vol. February, 2005.
 - [8] S. L. Howitt and D. Richards, "The Human Machine Interface for Airborne Control of UAVs," presented at 2nd AIAA "Unmanned Unlimited" Systems, Technologies, and Operations—Aerospace, Land, and Sea Conference and Workshop, San Diego, CA, 2003.
 - [9] M. L. Cummings and S. Guerlain, "Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles," *Human Factors*, in press.
 - [10] D. R. Olsen and S. B. Wood, "Fan-out: Measuring Human Control of Multiple Robots," presented at CHI2004, Vienna, Austria, 2004.
 - [11] D. R. Olsen and M. A. Goodrich, "Metrics for Evaluating Human-Robot Interactions," presented at Performance Metrics for Intelligent Systems, Gaithersburg, MD, 2003.
 - [12] M. L. Cummings, "The Need for Command and Control Instant Message Adaptive Interfaces: Lessons Learned from Tactical Tomahawk Human-in-the-Loop Simulations," *CyberPsychology and Behavior* vol. 7, 2004.
 - [13] M. L. Cummings and P. M. Mitchell, "Managing Multiple UAVs through a Timeline Display," presented at AIAA InfoTech, Arlington, VA, 2005.
 - [14] T. B. Sheridan and W. L. Verplank, "Human and Computer Control of Undersea Teleoperators," MIT, Cambridge, Man-Machine Systems Laboratory Report 1978.
 - [15] G. Osga, K. Van Orden, N. Campbell, D. Kellmeyer, and D. Lulue, "Design and Evaluation of WarfighterTask Support Methods in a Multi-Modal Watchstation," SPAWAR, San Diego 1874, 2002.
 - [16] M. L. Cummings and P. J. Mitchell, "Automated Scheduling Decision Support for Supervisory Control of Multiple UAVs," *Journal of Aerospace Computing, Information, and Communication*, in press.
 - [17] B. Hilburn, P. G. Jorna, E. A. Byrne, and R. Parasuraman, "The Effect of Adaptive Air Traffic Control (ATC) Decision Aiding on Controller Mental Workload," in *Human-automation Interaction: Research and Practice*. Mahwah, NJ: Lawrence Erlbaum, 1997, pp. 84-91.

- [18] M. L. Cummings, "Automation Bias in Intelligent Time Critical Decision Support Systems," presented at AIAA Intelligent Systems, Chicago, IL, 2004.