Modeling the Impact of Workload in Network Centric Supervisory Control Settings

M. L. Cummings & C. E. Nehme

Massachusetts Institute of Technology

Abstract: **The Department of Defense's vision of network centric operations will likely bring about higher operator mental workload due to the large volume of incoming information. As a result, it is critical that systems designers develop predictive models of both human and system performance, such that they can determine how a proposed technology will influence not only operator workload, but also system performance. To this end, this paper introduces a discrete event simulation approach to human-system modeling that includes a quantitative relationship between workload and performance, inspired by the Yerkes-Dodson relationship. Using the concept of utilization, or operator percent busy time, as a surrogate workload measure, we demonstrate that a quantitative instantiation of an inverted-U workload-performance curve improves discrete event simulation model predictions in a human supervisory control model of single operator control of multiple unmanned vehicles. While this work generates the first known empirical evidence of a parabolic workload-performance as a variable in a quantitative predictive human performance model, this effort is preliminary and future research implications are discussed.**

***Index Terms*—Network centric operations, utilization, discrete event simulation, human performance modeling, unmanned vehicles.**

I. Introduction

Network Centric Operations (NCO), also known as Network Centric Warfare (NCW), is a concept of operations envisioned to increase combat power by electronically linking or networking relevant entities across a battle field. As defined by the Department of Defense (DoD), key components of NCO include information sharing and collaboration which will theoretically promote shared situational awareness and overall mission success [1]. What this practically means is that personnel will have access to exponential amounts of information as compared to today's forces, and thus the information intake for the average NCO operator will be higher than ever before in command and control settings. As a result, mental workload will likely increase, and a large body of human factors literature has demonstrated that as mental workload increases, performance can be negatively affected.

Given that NCO will likely bring about higher mental workload due to the large volume of incoming information, it is critical that systems designers develop predictive models of both human and system performance such that they can determine given a particular concept of operations, or a proposed technology, what the impact will be on operator workload, particularly in terms of reaching critical performance thresholds. Towards this end, this paper will introduce a new discrete-event simulation approach to human performance modeling that includes a quantitative relationship between workload and performance.

II. Background

To achieve the vision of NCO, the DOD will increasingly need to rely upon the concept of supervisory control, which is when an operator intermittently interacts with a computer that closes an autonomous control loop [2]. In particular, there is currently a significant thrust towards including unmanned vehicles in military operations, and human interaction with such vehicles, either on the ground or in the air, is inherently a supervisory control problem. Moreover, given increased autonomy of unmanned vehicles (UVs), a human operator's role is shifting from controlling one vehicle in a team of operators to supervising multiple vehicles individually [3]. However, a primary limiting factor in this one operator-many UV vision is operator workload. Indeed, this limitation on the control of multiple UVs extends to any supervisory control tasks requiring divided attention across multiple tasks such as air traffic control and even supervisors multi-tasking in a command center such as an air operation center.

Mental workload results from the demands a task imposes on the operator's limited resources; it is fundamentally determined by the relationship between resource supply and task demand [11]. While there are a number of different ways to measure workload [4-6], given the temporal nature of supervisory control systems, particular those in command and control settings, we propose that the concept of utilization is an effective proxy for measuring mental workload. Utilization is a term found in systems engineering settings and refers to the "percent busy time" of an operator, i.e., given a time period, what percentage of time was that person busy. In supervisory control settings, this is generally meant as the time an operator is directed by external events to complete a task (e.g., replanning the path of an unmanned aerial vehicle (UAV) because of an emergent target). For example, given an operator shift of 8 hours, the utilization would be that amount of time the operator was engaged in directed tasks divided by 480 minutes. What is not included in this measurement is the time spent monitoring the system, i.e., just watching the displays and/or waiting for something to happen.

The concept of utilization as a mental workload measure has been used in numerous studies examining supervisory controller performance [7-9], and these studies generally support that when tasked beyond 70% utilization, operators' performances decline. While arguably this is not a perfect measure of mental workload, one strength of such a measure is its ratio scale, which allows it to be used in quantitative models. Given the previous research showing that supervisory control performance drops when utilization is greater than 70%, we investigated whether using such a numerical relationship could be used to not just describe observed human behavior, but also be used to predict it.

In order to investigate this possible relationship between workload (as measured by utilization) and performance, we elected to use a discrete event simulation (DES) model. A DES model can represent a system as it evolves over time by representation of events, i.e., instantaneous occurrences that may change the state of the system [10], and is particularly suited to model supervisory control systems due to their time-critical, event-driven nature which are characteristic of NCO environments. Discrete event simulations are based on queuing theory, which model the human as a single server serially attending the arrival of events [11-13]. These models can also be extended to represent operator parallel processing through the introduction of multiple servers [14, 15]. The next section presents an overview of the DES model and the inclusion of the utilization-performance relationship.

III. The Model

In our model, the NCO human operator is responsible for multiple tasks, specifically the management of multiple unmanned vehicles. This operator is modeled as a server in a queuing system with discrete events representing both endogenous and exogenous situations an operator must address. Endogenous events, which are vehicle-generated or operator induced, are events created internally within the supervisory control system, such as when an operator elects to re-plan an existing UV path in order to reach a goal in a shorter time. Events which result from unexpected external environmental conditions that create the need for operator interaction are defined as exogenous events, such as emergent threat areas which require re-planning vehicle trajectories. All supervisory control systems have such events arriving into the system, often of differing priorities which will subsequently be addressed.

The design variables that serve as inputs to the model in Figure 1 are composed of variables related to the vehicle team (number of vehicles and the vehicle types, as expressed through neglect time), the human operator (interaction

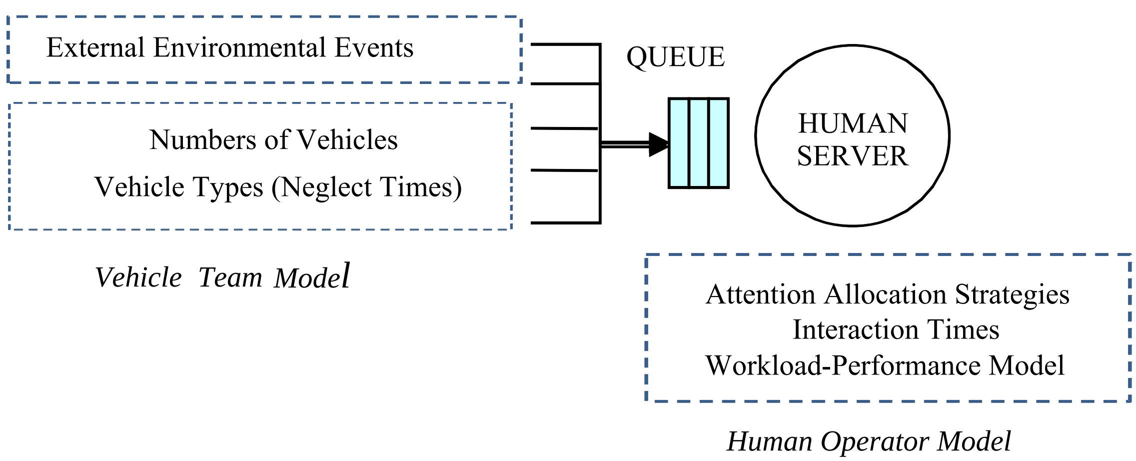


Figure 1: A high level representation of the discrete event simulation model

times, operator attention allocation strategies, and the operator situational awareness), and a model of environment unpredictability. These are discussed below in further detail.

A. Vehicle Team Input Variables

In our model in Figure 1, we include the number and type of vehicles in the system. While the number of vehicles directly affects the number of events arriving to the operator's queue, we also need to represent the varying types of vehicles, since in the future NCO vision, one person can control multiple types of UVs. We do this through the concept of neglect time (NT). NTs represent the time a vehicle can operate without human intervention, and different NTs represent different vehicle types.

B. Human Operator Input Variables

The length of time it takes the operator to service an event, also known as interaction time, is captured through a distribution of event service times; thus, the service process can be described by a probabilistic distribution representing the interval from the time the operator decides to service a particular event until the operator is done servicing. We model the operator's strategies for handling priorities, known as attention allocation [16], through the queuing policy, i.e., which task waiting in a queue the operator elects to service. Examples include the first-in-first-out (FIFO) queuing scheme as well as the highest attribute first (HAF) strategy [17]. The HAF strategy assumes that high priority events are serviced first, a typical strategy employed in high risk command and control settings. We also model the operator's management strategy, which is the willingness of the operator to allow the vehicles to operate on their own without intervention. This variable can be seen as a trust proxy since those operators who do not trust the system will intervene more frequently than those who do.

C. The Workload-Performance Model

As illustrated in a previous study examining single operator control of multiple unmanned vehicles [18], successful models of human performance in such systems should contain some representation of the inherent delays that humans will introduce into these complex systems. These delays occur through queues that build for multiple tasks since humans cannot instantaneously make decisions, or through delays due to a loss of situation awareness because operators do not realize the system needs their attention. While the delays due to increasing queue size are accounted for in a typical discrete even simulation model, the impact of the loss of situation awareness on a system is far more difficult to both measure and predict. As seen in Figure 1, we propose that we can account for the delays in the system due to this loss of operator situation awareness through a workload-performance model.

This workload-performance model is inspired by the Yerkes-Dodson inverted-U relationship (often called a law, which has been a point of contention [19]), notionally illustrated in Figure 2. For our purposes, we will use the previously discussed measure of utilization as our proxy measure for workload. While the original Yerkes-Dodson research focused on stimulus strength and learning [20], a similar effect on the arousal level and performance was found in later work [21]. In general, this relationship states that people work best under moderate levels of workload, and that high or low levels of workload will result in degraded operator performance.

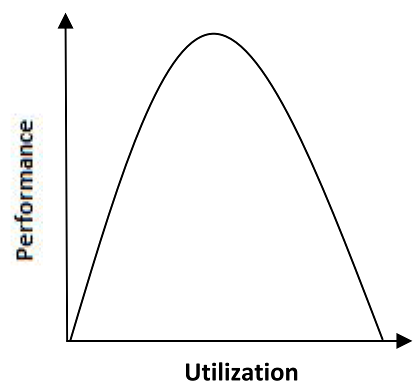


Figure 2. The notional workload-performance relationship

The Yerkes-Dodson relationship has been labeled as overly simplistic and more of an appeal to common sense that to scientific rigor [19], which are no doubt true to some degree. However, previous research has shown direct evidence of such a relationship in supervisory control performance using utilization as a workload measure [8]. Moreover, when attempting to build models of any system's performance, human or otherwise, Occam's Razor prevails, and if a relatively simple relationship can provide robust predictive results across a number of independent data sets when embodied in a human-system model, then this relationship should not be dismissed, but rather more deeply investigated.

Operator utilization, our measure of arousal, is hypothesized to ultimately affect performance, such that it is degraded at both high and low ends of the utilization curve (Figure 2). When operators are under high levels of utilization, it is expected that they can be too busy to accumulate the information that is required to perform at optimal levels. When operators are under-utilized, we propose that due to a low level of arousal and complacency, they can overlook information from the environment or engage in a non-related task (such as reading the paper), which also leads to degraded performance. There is a possible connection between situation awareness and this workload-performance curve, which has been discussed in detail elsewhere [22, 23].

IV. Experimental Validation for the Workload-Performance Model

In order to validate the hypothesis that the previously described workload-performance model was necessary in the discrete event simulation (DES) model illustrated in Figure 1, experimental data from a multiple unmanned vehicle simulation was analyzed. Given the previous debate as to the validity of the inverted-U Yerkes Dodson relationship, it was critical to test to see whether the inclusion of the proposed parabolic relationship data (Figure 2) improved the DES model's ability to replicate observed human experimental data.

A. Experimental Apparatus and Participants

In this experimental test bed, the operator's mission was to supervise multiple unmanned ground vehicles via a supervisory control interface in order to remove as many objects as possible from a maze in an 8-minute time period. The objects were randomly spread through the maze, which was initially unknown. However, as each UV moved about the maze, it created a map which it shared with the subject and the other UVs in the team. There were 22 possible objects to collect during a session. Vehicles contained some level of automatic path planning as well as well as automated goal designation. However, these algorithms were sometimes sub-optimal, which could cause the vehicles to become stuck, thus requiring operator intervention. The typical of heuristic algorithms, they also were sometimes inefficient in goal tasking, which could be significantly improved by human intervention. These kinds of suboptimal automation behaviors are very common in systems with embedded autonomous planning, as these algorithms are inherently brittle and unable to respond to complex system events [24].

The human-UV supervisory control interface was the two-screen display shown in Figure 3. On the left screen, the map of the maze was displayed, along with the positions of the UVs and objects with known locations. This screen also indicated to operators which vehicle was in a particular state and which needed intervention. It also included a time line that indicated mission temporal progress. In order to remove an object from the maze, operators were required to first complete a visual task, which is representative of real world intelligence, surveillance, and reconnaissance missions. To simulate this visual task, users were asked, using the right screen in Figure 3, to identify a city on a map of the mainland United States using *Google Earth*-style software.

Sixteen participants between the ages of 18 and 45 participated in this study. The experimental results reported here represent a subset of experimental results from larger studies investigating predictive modeling techniques for supervisory control systems [16, 25, 26]. Participants completed three comprehensive practice sessions, followed by four test sessions with 2, 4, 6, and 8 vehicle team sizes (randomized and counterbalanced to control for any learning effect. Participants were paid $10 per hour; the highest scorer also received a $100 gift certificate.

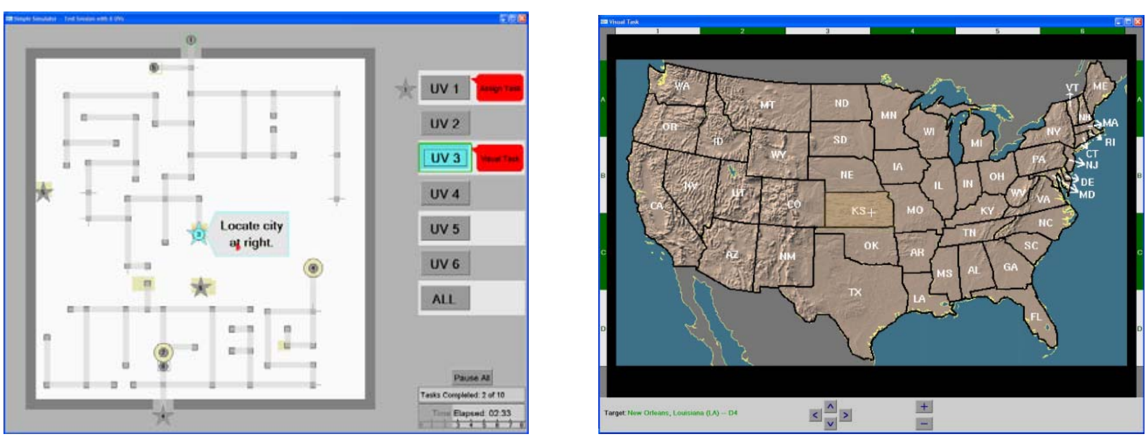


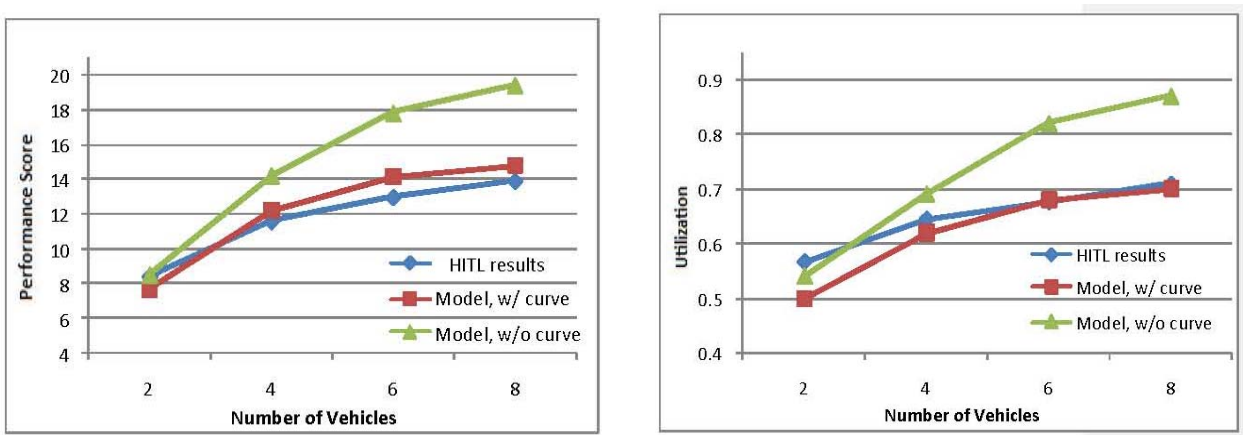
Figure 3. Two-screen interface for single operator control of multiple UVs

B. Results

In order to compare the DES model results to the experimental results generated from the experiment, probabilistic distributions were obtained for the arrival and service rates as experienced by the participants in the study. These distributions included the service times of the visual tasks as well as the replanning of vehicles between different objects. The inter-arrival time distributions representing those times between autonomous replanning events were also included. In terms of performance, the DES model awarded a point for the servicing of every vehicle-generated event, e.g., replanning a vehicle when it became stuck. Using the distributions of arrival rates and service times generated from the data in the experiment, 10,000 trials were conducted with the DES, in order to compare the results with the human-in-the-loop experiment. The workload-performance model was a parabolic curve which penalized operators by essentially adding delay time to their interactions based on the utilization measure. This, in turn, slowed the rate at which participants acquired points. Further details about both the simulation test bed and the model can be found in [27].

The observed system performance and operator utilization from the user study are compared with the model's estimates in Figure 4. Two dependent variables were measured, the performance (a combination of the objects collected and the correction of erroneous automation behavior), and utilization (as defined previously, the time operators were engaged in tasks divided but the total experimental session time.) In Figure 4, the results of the human-in-the-loop (HITL) study can be seen in terms of performance and utilization, as well as the results of the model if the workload-performance curve is or is not included.

Without the workload-performance curve, the model tended to over-predict performance for the 4, 6 and 8 vehicle conditions and under predict for the 2 vehicle condition. Including the workload-performance curve in the DES model led to decreased performance scores (and thus improved predictions), except in the 2 vehicle condition where the model was already slightly under-predicting performance. However, the model's results for system performance



a. Performance Score **b. Utilization**

Figure 4. Human-in-the-loop (HITL) results versus model predictions

when including the workload-performance curve score were all within the 95% confidence intervals, except for the 2 vehicle case. The model without the workload-performance curve was well outside the 95% confidence interval in all cases.

The inclusion of the workload-performance curve in the model also had an overall positive effect in terms of utilization predictions, as seen in Figure 4b. Without this curve, the DES model tended to over-predict operator utilization for the 4, 6 and 8 vehicle conditions and under-predict for the 2 vehicle condition. When the operator is processing too many tasks, utilization is high which can lead to low performance (as computed by the model). Also, as suggested by Yerkes-Dodson, and confirmed in our empirical results, low utilization translates into fewer correct tasks processed by the operator, thus influencing performance as seen in Figure 4a.

An additional observation about the accuracy of the model can be made from these results, in that the model is more accurate for larger teams than small teams. This trend appears to be caused, at least to some degree, by overly high penalties associated with low utilization in the workload-performance curve in the model. This model parameter leads to the under estimation of both operator utilization and system performance in this condition. This suggests that while the concept of high or low workload contributes to lower operator, and thus system, performance, the symmetric relationship as depicted in Figure 2 inspired by Yerkes-Dodson, is not quite accurate. This observation was seen again in a second related experiment, detailed in the next section.

V. Additional Experimental Validation for the Workload-Performance Relationship

Given the results in the previous section, additional experimental data was gathered to assess the nature of the workload-performance curve relationship as depicted in Figure 2. The previous section confirmed that including such a relationship in a DES model significantly improved the predictive ability of this model for actual human performance. However, there was an indication that the symmetrical shape of the curve was not accurate, causing the DES model to under-predict utilization and performance. Using data gathered from a completely independent experiment from the one previously discussed, we investigated whether the model, including the workload-performance curve, would also accurately replicate observed HITL performance. In addition, given the previous results, we also wanted investigate the observed shape of the performance and utilization curves. We could not do this in the previous experiment since the experiment was not explicitly designed to investigate this curve, rather, we hypothesized post-hoc that such a relationship was needed after seeing the results.

A. Experimental Test-bed

The Research Environment for Supervisory Control of Heterogeneous Unmanned Vehicles (RESCHU) simulation test bed allows operators to control a team of UVs composed of unmanned air and underwater vehicles (UAVs and UUVs). All vehicles engage in surveillance tasks, with the ultimate mission of locating specific objects of interest in urban coastal and inland settings. There is a single UUV type, and two UAV types. The first is a high altitude long endurance (HALE) UAV that provides high-level sensor coverage (akin to a Global Hawk UAV) for target identification, while the other, a medium altitude long endurance (MALE) UAV, provides more low-level target surveillance and video gathering (similar to a Predator UAV). Thus, an operator can control up to three different vehicle types. Basic flying and navigation tasks for the different vehicles is highly automated with embedded autonomy to allow operators to concentrate on payload tasking [28].

The RESCHU interface consists of five major sections (Figure 5). The map displays the locations of vehicles (U-shaped icons), threat areas (circles), and areas of interests (AOIs, indicated by diamonds). Vehicle control is carried out on the map, such as changing vehicle paths, adding a waypoint (a destination along the path), or assigning an AOI to a vehicle. The main events in the mission (i.e., vehicles arriving to goals, or automatic assignment to new targets) are displayed in the message box, along with a timestamp. When the vehicles reach an AOI, a simulated video feed is displayed in the camera window, and the operator then visually identifies a target in this simulated video feed for just MALE UAVs and UUVs. Example targets and objects of interest include cars, swimming pools, helipads, etc. When vehicles complete their assigned visual tasking, an automated-path planner automatically assigns the HALE UAV to an AOI that needs intelligence, and the MALE UAVs and UUVs to AOIs to pre-determined targets. As in the previous experiment, the automatically-assigned AOIs are not necessarily the optimal choice. As a consequence, an operator can change the assigned AOI, and avoid threat areas by changing a vehicle's

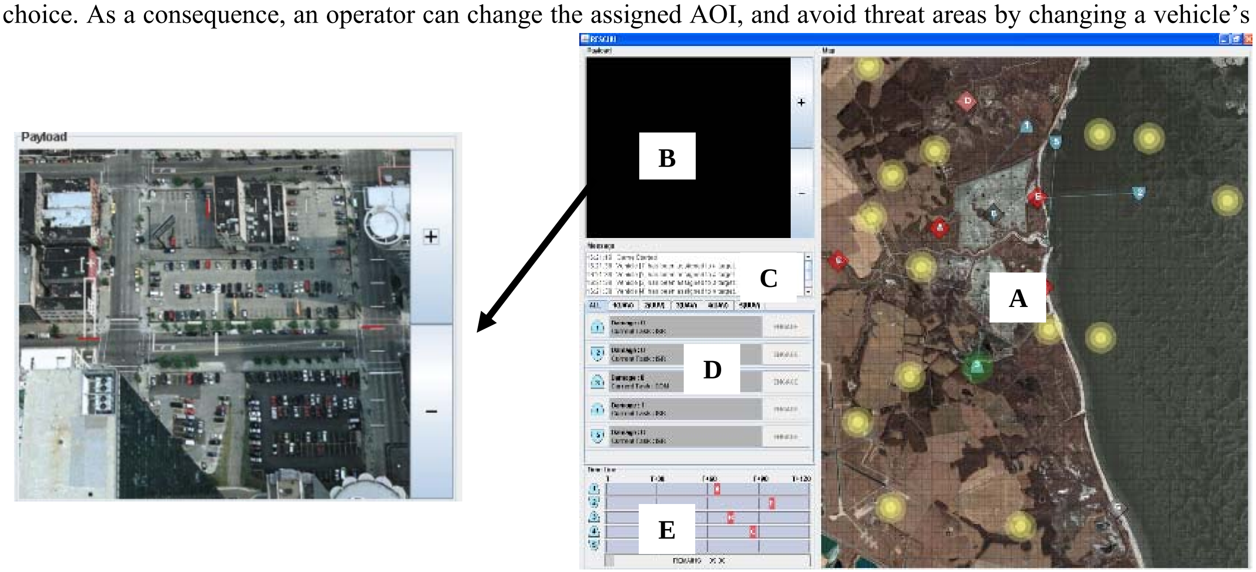
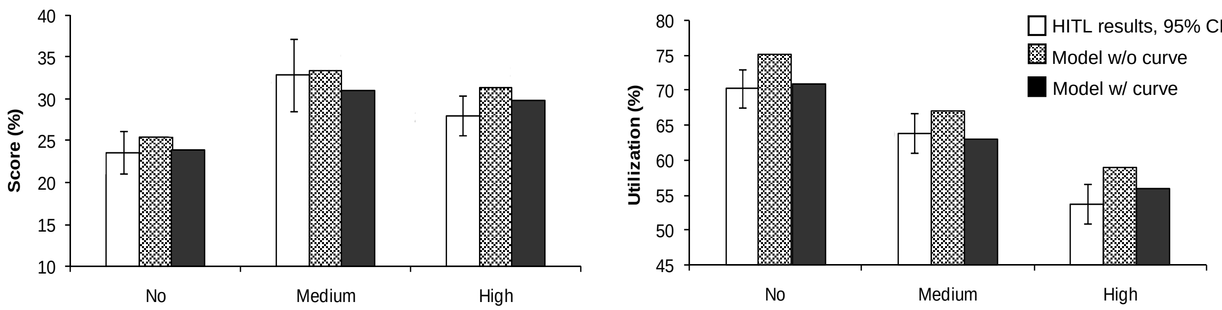


Fig 5. RESCHU interface (A: map, B: camera window, C: message box, D: control panel, E: timeline)

goal or adding a waypoint to the path of the vehicle in order to go around the threat area.

Lastly, the control panel provides vehicle health information, as well as information on the vehicle's mission. The timeline displays the estimated time of arrival to waypoints and AOIs. Beneath the timeline is a mission progress bar that shows the amount of time remaining in the total simulation. Conceptually, the two interfaces depicted in Figures 3 and 5 are the same, however, the RESCHU interface is a much more realistic simulation environment than the one



b. Performance Score b. Utilization

Figure 6. Human-in-the-loop (HITL) results versus model predictions per level of heterogeneity

used in the previous experiment, and more directly tied to the DoD vision of one operator controlling multiple unmanned vehicles of various types.

B. Participants and Procedure

Given the relatively small numbers of participants in the previous experiment, we wanted to ensure that any significant trends resulting from this experiment had higher statistical power. As a result, seventy-four participants, six females and 68 males, between the ages of 18-50 completed the experiment. The participant who scored the highest received a $200 gift certificate. The experiment was conducted online with an interactive tutorial and an open-ended practice session. The website was password protected and participation was via invitation. All data were recorded to an online database. Demographic information was collected via a questionnaire presented before the tutorial. After participants felt comfortable, they could end the practice session and start the ten minute experimental session.

Participants were instructed to maximize their score by 1) avoiding dynamic threat areas, 2) completing the maximum number of search tasks correctly, 3) minimizing vehicle travel times between AOIs, 4) ensuring a vehicle was always assigned to an AOI whenever possible. They controlled five vehicles each, of mixed vehicle types, which will be detailed in the next section.

C. Experimental Design and Independent Variables

The experiment was a completely randomized design with vehicle team heterogeneity level as a between-subject condition: none (n=26), medium (n=25), and high (n=23). The no heterogeneity condition composed of five MALE UAVs. The medium heterogeneous level had three MALE UAVs and two UUVs. Because the UUVs were slower than UUVs, they produced events less frequently. HALE UAVs visited targets that required identification, and once identified, either MALE UAVs or UUVs were assigned to the targets. Because the UUVs were slower than UAVs and the HALE UAVs did not have an associated visual task, the no heterogeneity condition composed of five MALE UAVs was the highest tempo scenario, (meaning events arrived more frequently) followed by the medium and then the high heterogeneity conditions. Such variations in team composition helped to determine how robust the model is, as the vehicles in the first experiment were a homogeneous set.

As in the previous experiment, the variables of interest for evaluating model predictions were score and operator utilization. The performance score was calculated as the proportion of the number of targets correctly identified normalized by the number of all possible targets that could have been identified. Operator utilization was calculated as in the previous experiment.

VI. Results & Discussion

The observed average performance scores and utilizations for all three vehicle heterogeneity levels are presented in Figure 6, along with the predictions of the model with and without the workload-performance curve. The curve used to penalize operators by adding delay times for high and low utilizations was based on the distributions derived from the experimental data. As seen in the previous experiment, the inclusion of the inverted U-shaped workload-performance curve allows the model's predictions to fall within the 95% confidence interval across all three different combinations of vehicle types. The DES model without the curve consistently over-predicted both performance scores and utilizations, and thus was significantly improved with the addition of the workload-performance model. These results, in combination with the previous experimental results, provide clear evidence that a parabolic workload-performance curve is critical in DES modeling efforts for human operator interaction with supervisory control systems.

Regarding the shape of the inverted U curve seen in the human-in-the-loop experiment, the distribution of the performance scores in 10% intervals (Figure 7) across all aggregated trials mirrors the hypothesized inverted-U relationship such that over or under utilization caused mission performance to degrade. Pair-wise comparisons for Figure 7 revealed that 60-70% utilization corresponded to significantly higher scores than the majority of other utilization values ($p \leq .02$ across the various categories). In addition, 80-90% utilization resulted in lower performance scores than both 70-80% ($p=.03$) and 40-50%, ($p=.04$) utilization. Previous studies have also shown that when the operators work beyond 70% utilization, their performance degrades significantly [7, 29], so these results confirm these previous results that there can be a threshold for performance in terms of operator utilization.

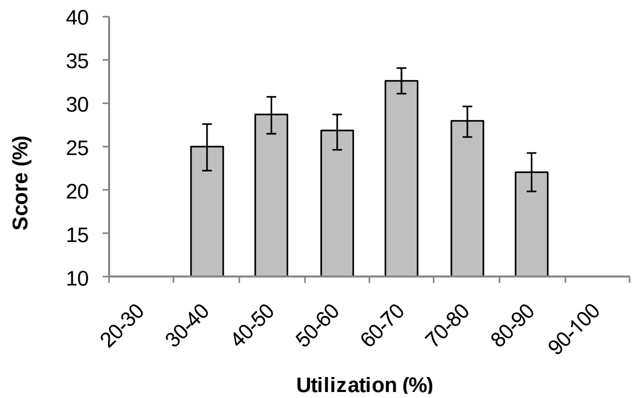


Figure 7. Experimental results for score vs. utilization (with standard error bars)

Given that the performance score (the percentage of possible targets acquired) was context-specific, another dependent variable that is both more generalizable and more indicative of the possible negative impacts of performance problems, is the delay time caused by operator inattention. This was discussed previously as delay times induced into the system by the loss of operator situation awareness, and these are the source of delay penalties described earlier. Previous research has shown that these delays caused by operator inattention can account for the largest part of vehicle wait time, and significantly reduce the overall number of vehicles that a single operator could control [18]. Since delay times caused by an operator who has lost situation awareness are a more objective measure of performance that a derived performance score, we examined how these wait times mapped to operator utilizations.

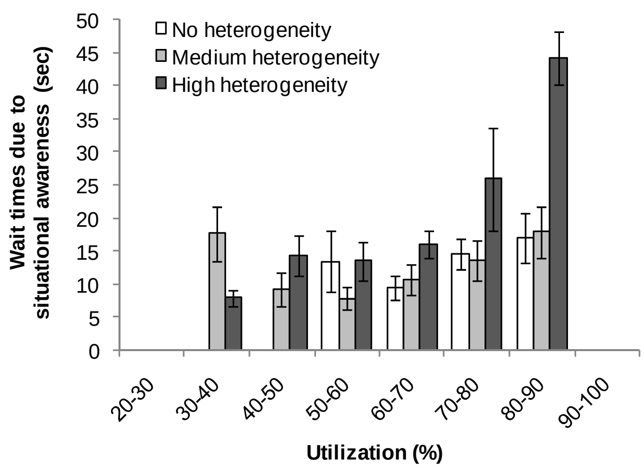


Figure 8. The effects of increasing utilizations for induced delay times across the three levels of UV team heterogeneity

When these wait times caused by operator loss of situation awareness are mapped to the 10% utilization bins (Figure 8) (causing a U shaped curve due to negative consequences of wait times as opposed to the inverted-U positive outcome of performance score), the results differ somewhat from those in Figure 7. In the case of those operators responsible for the 2 UUVs, 2 MALE UAVs, and 1 HALE UAV (high heterogeneity), there is a clear skew to the higher utilization bins. This UV team was more cognitively complex to control than the no homogeneity team, which as seen in Figure 8, produced a workload-performance curve more in keeping with the notional U-shape function (inverted) in Figure 2. The medium level of heterogeneity operators experienced slight skews to the higher utilization bins, which suggests that for this experiment, as the complexity of the task increased, workload also increased.

VII. CONCLUSIONS

With the voluminous increase in incoming information in network centric operations, identifying plateaus of optimal performance as a function of workload, as well as critical thresholds for degraded performance are key to designing systems that operators can effectively manage. Predicting regions of optimal performance or points of cognitive saturation is difficult for NCO systems because they are inherently task dependent, with potential varying levels of automation, operational tempos, training, experience, etc. Thus it is critical to develop simulation models for these futuristic systems that allow for exploration of these different design parameters without conducting extensive experimentation. However, these simulation models need accurate representations of human performance. In this paper, we have demonstrated that such a model is achievable through a discrete event simulation model, but that representing the negative impact of operator performance as a function of workload in a parabolic form is pivotal in improving this model's predictions.

While this work has generated the first known empirical evidence that some form of a parabolic workload-performance curve, suggested by the Yerkes-Dodson relationship, is both observable and when replicated in a quantitative model, adds significant value in human supervisory control discrete event simulation models of unmanned vehicles, this effort should still be considered preliminary. There are a number of related issues that require further investigation.

First, as suggested by Figure 8 and the results of the first experiment, true workload-performance curves are likely not parabolic in the symmetrical sense. As seen in the second experiment, increasingly complex mixtures of vehicle teams caused different overall utilizations and correlated performances. Thus the nature of the task could dramatically vary an accurate workload-performance representation. However, in any modeling effort, it is not possible to always capture every variable influencing a system, thus if more general workload-performance curves could be found for task types, these relationships could be very valuable to human-system modeling efforts.

Another issue is the temporal and dynamic nature of utilizations. All utilizations reported here were post-hoc aggregate measures, which are not accurate for any instantaneous measure of workload. Thus more research is need to determine how utilizations can be measured in an online fashion, and moreover, what thresholds truly indicate poor performance, i.e., does a person need to work at or above 70% utilization for some period of time before the negative effects are seen in human and/or system performance.

Another related area that requires further investigation is the rate of onset of high workload or utilization periods. The true measure of the impact of workload on performance may not be sustained utilization, but rather the onset rates of increasing utilizations. The other side of the curve represents cognitive under-load, and the nature of sustained and variable rates of utilization changes also deserves further scrutiny.

Such investigations will be crucial in both aiding supervisory control modeling efforts, but they are also potentially valuable in the field of dynamic, adaptive automation design. If successful performance models of over or under cognitive load based on utilization can be developed, then more reliable forms of adaptive automation can be developed such that automation can intervene or assist human operators when a transition into a negative workload-performance region occurs.

Lastly, more fundamental work is needed to determine the nature of underlying physiologic or cognitive and psychological factors in contributing to performance decrements when operators move beyond some plateau. Previous research by Hancock and Warm [30] suggests that operators seek a zone of maximal psychological capability that is constrained by their physiologic capability, and that once these zones are exceeded, performance degradation can occur. More work is needed to determine how psychological or physiologic processes are similar or different under high and low utilizations. Furthermore, future efforts are needed to determine if these zones can be generalized in such a way to include them in a quantitative model, in order to replicate observed human behavior so that the models can be used to predict both human and system performance.

Acknowledgment

This research was sponsored by the Office of Naval Research, Charles River Analytics, Inc., and Lincoln Laboratory. Special thanks to Dr. Birsen Donmez for the statistical support and Yale Song and Dr. Jacob Crandall for the test bed support.

References

[1] DoD, "Network Centric Warfare: Department of Defense Report to Congress," Office of the Secretary of Defense, Washington DC2001.

[2] T. B. Sheridan, *Telerobotics, automation, and human supervisory control*. Cambridge, MA: The MIT Press, 1992.

[3] DOD, "Unmanned systems roadmap," Office of the Secretary of Defense, Washington, D.C. 2007-2032, 2007.

[4] P. E. Pina, M. L. Cummings, J. W. Crandall, and M. Della Penna, "Identifying Generalizable Metric Classes to Evaluate Human-Robot Teams," in *Metrics for Human-Robot Interaction Workshop at the 3rd Annual Conference on Human-Robot Interaction*, Amsterdam, Netherlands, 2008.

[5] C. D. Wickens and J. G. Hollands, "Engineering Psychology and Human Performance," 3rd ed Upper Saddle River, NJ: Prentice-Hall, 2000.

[6] W. W. Wierwille and F. T. Eggemeier, "Recommendations for Mental Workload Measurement in a Test and Evaluation Environment," *Human Factors,* vol. 35, pp. 263-281, 1993.

[7] D. K. Schmidt, "A Queuing Analysis of the Air Traffic Controller's Workload," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 8, pp. 492-498, 1978.
[8] M. L. Cummings and S. Guerlain, "Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles," *Human Factors,* vol. 49, pp. 1-15, 2007.
[9] W. B. Rouse, *Systems Engineering Models of Human-Machine Interaction*. New York: North Holland, 1983.
[10] A. M. Law and W. D. Kelton, *Simulation modeling and analysis*, 3rd ed. Boston: McGraw-Hill International Series, 2000.
[11] J. W. Senders, "The human operator as a monitor and controller of multidegree of freedom systems," *IEEE Transactions on Human Factors in Electronics,* vol. 5, pp. 2-5, 1964.
[12] T. B. Sheridan, "On how often the supervisor should sample," *IEEE Transactions on Systems, Science, and Cybernetics,* vol. 6, pp. 140-145, 1969.
[13] J. R. Carbonell, J. L. Ward, and J. W. Senders, "A queuing model of visual sampling experimental validation," *IEEE Transactions on Man Machine Systems,* vol. 9, pp. 82-87, 1968.
[14] Y. Liu, "Queuing network modeling of human performance of concurrent spatial and verbal tasks," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 27, pp. 195-207, 1997.
[15] Y. Liu, R. Feyen, and O. Tshimhoni, "Queuing network-model human processor (QN-MHP): a computational architecture for multitask performance in human-machine systems," *ACM Transactions on Computer-Human Interaction,* vol. 13, pp. 37-70, 2006.
[16] J. W. Crandall and M. L. Cummings, "Identifying predictive metrics for supervisory control of multiple robots," *IEEE Transactions on Robotics - Special Issue on Human-Robot Interaction,* vol. 23, pp. 942-951, 2007.
[17] M. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 2002.
[18] M. L. Cummings and P. J. Mitchell, "Predicting controller capacity in supervisory control of multiple UAVs," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* vol. 38, pp. 451-460, 2008.
[19] P. A. Hancock and H. C. N. Ganey, "From the inverted-U to the extended-U:The evolution of a law of psychology. ," *Human Performance in Extreme Environments,* vol. 7, pp. 5-14, 2003.
[20] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology,* vol. 18, 1908.
[21] D. O. Hebb, "Drives and the C. N. S. (conceptual nervous system)," *Psychological Review,* vol. 62, pp. 243-254, 1955.
[22] C. Nehme, J. W. Crandall, and M. L. Cummings, "Using Discrete Event Simulation to Model Situational Awareness of Unmanned-Vehicle Operators," in *Modeling, Analysis and Simulation Center Capstone Conference*, Norfolk, VA, 2008.
[23] B. Donmez, C. Nehme, M. L. Cummings, and P. de Jong, "Modeling Situational Awareness in a Multiple Unmanned Vehicle Supervisory Control Discrete Event Simulation," *IEEE Systems, Man, and Cybernetics,Part A Systems and Humans,* in review.
[24] P. Smith, E. McCoy, and C. Layton, "Brittleness in the design of cooperative problem-solving systems: The effects on user performance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* vol. 27, pp. 360-371, 1997.
[25] J. W. Crandall and M. L. Cummings, "Attention Allocation Efficiency in Human-UV Teams," in *AIAA Infotech@Aerospace Conference* Rohnert Park, CA, 2007.
[26] J. Crandall and M. L. Cummings, "Developing Performance Metrics for the Supervisory Control of Multiple Robots," in *Human Robotics Interaction 2007*, Washington DC, 2007.
[27] C. E. Nehme, "Modeling human supervisory control in heterogeneous unmanned vehicle systems," in *Aeronautics and Astronautics*. vol. Doctor of Philosophy Cambridge, MA: Massachusetts Institute of Technology, 2009.
[28] M. L. Cummings, S. Bruni, S. Mercier, and P. J. Mitchell, "Automation architecture for single operator, multiple UAV command and control," *The international Command and Control Journal,* vol. 1, pp. 1-24, 2007.
[29] M. L. Cummings and S. Guerlain, "Developing operator capacity estimates for supervisory control of autonomous vehicles," *Human Factors,* vol. 49, pp. 1-15, 2007.
[30] P. A. Hancock and J. S. Warm, "A Dynamic Model of Stress and Sustained Attention," *Human Factors,* vol. 31, pp. 519-537, 1989.