

MIT Open Access Articles

Modeling workload impact in multiple unmanned vehicle supervisory control

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Donmez, B., C. Nehme, and M.L. Cummings. "Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control." *Systems, Man and Cybernetics, Part A: Systems and Humans*, IEEE Transactions On 40.6 (2010) : 1180-1190. © 2010 IEEE.

As Published: <http://dx.doi.org/10.1109/TSMCA.2010.2046731>

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: <http://hdl.handle.net/1721.1/65349>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control

Birsen Donmez, *Member, IEEE*, Carl Nehme, and Mary L. Cummings, *Senior Member, IEEE*

Abstract—Discrete-event simulations for futuristic unmanned vehicle (UV) systems enable a cost- and time-effective methodology for evaluating various autonomy and human-automation design parameters. Operator mental workload is an important factor to consider in such models. We suggest that the effects of operator workload on system performance can be modeled in such a simulation environment through a quantitative relation between operator attention and utilization, i.e., operator busy time used as a surrogate real-time workload measure. To validate our model, a heterogeneous UV simulation experiment was conducted with 74 participants. Performance-based measures of attention switching delays were incorporated in the discrete-event simulation model by UV wait times due to operator attention inefficiencies (WTAIs). Experimental results showed that WTAI is significantly associated with operator utilization (UT) such that high UT levels correspond to higher wait times. The inclusion of this empirical UT-WTAI relation in the discrete-event simulation model of multiple UV supervisory control resulted in more accurate replications of data, as well as more accurate predictions for alternative UV team structures. These results have implications for the design of future human-UV systems, as well as more general multiple task supervisory control models.

Index Terms—Attention allocation, operator utilization (UT), queuing theory, simulation, unmanned vehicles (UVs).

I. INTRODUCTION

SUPERVISORY control refers to intermittent operator interaction with a computer that closes an autonomous control loop [1]. With increased autonomy of unmanned vehicles (UVs), a human operator's role shifts from controlling one vehicle to supervising multiple vehicles [2]. In the future, it is likely that a team of UVs will be composed of vehicles that vary in their capabilities or their assigned tasks, resulting in a heterogeneous system [3], [4]. Although the appropriate size of a team is mission dependent, several experiments have all reached the same conclusion: There exists some upper bound to the number of vehicles that can be supervised by a

single operator [5], [6]. To determine appropriate UV system architectures, it is critical to understand the impact of varying system design variables, such as level of vehicle autonomy, on the efficiency of human supervisory control of multiple UVs.

Human supervisory control is a complex system phenomenon with high levels of uncertainty, time pressure, and a dynamically changing environment. Discrete-event simulation (DES), which models a system as a chronological sequence of events representing changes in system states [7], is particularly suited to model supervisory control systems due to their time-critical event-driven nature. Such simulation models for futuristic systems allow for cost- and time-effective evaluation of different design parameters without conducting extensive experimentation, which is particularly critical in early conceptual design phases. Although other modeling techniques, including agent-based models [8], [9] can potentially be used to capture human-UV interactions, a DES model was chosen as the first step due to its ability to capture the temporal aspects of human-UV interactions. These temporal aspects determine important system limitations and are defined in the following sections.

Using DES-based approaches, a few studies have attempted to computationally predict operator capacity when controlling multiple UVs [10]–[12], which generally focused on the use of neglect time (NT) and interaction time (IT) to represent event arrival and service rate distributions. NT corresponds to the time that a robot or UV can be ignored before its performance drops below a predetermined threshold. IT is defined as the amount of time that the operator has to spend to bring a robot back to its peak performance. Although these previous studies focused on clearly observable state transitions, the inherent delays that humans introduce in supervisory control systems were not considered. Vehicle wait times due to attention inefficiencies (WTAIs) will occur as the operators fail to notice that the system needs their attention and have been shown to significantly affect system performance [13].

This paper uses experimental data to demonstrate the need to incorporate human attention inefficiencies in models of human-UV systems, as well as a methodology to do so. As a first step, the effects of mental workload on UV operator attention inefficiencies are investigated. The relation between operator utilization (UT), a surrogate measure of workload, and a performance-based measure of inattention is incorporated into a DES model of multiple UV supervisory control through WTAI. Without the inclusion of the UT-WTAI relation, the DES model fails to provide accurate replication and prediction of the observed data.

Manuscript received November 26, 2008; revised July 22, 2009. Date of publication June 14, 2010; date of current version October 15, 2010. This work was supported in part by the Charles River Analytics and by the Office of Naval Research. This paper was recommended by Associate Editor Y. Liu.

B. Donmez is with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada (e-mail: donmez@mie.utoronto.ca).

C. Nehme is with McKinsey and Company, Dubai, United Arab Emirates (e-mail: nehme@mit.edu).

M. L. Cummings is with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: missyc@mit.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSMCA.2010.2046731

II. BACKGROUND

DESS, which model the human as a single server that serially attends to the arrival of events, are based on the queuing theory [14]–[16]. These models can also be extended to represent operator parallel processing through the introduction of multiple servers [17], [18]. In addition to the aforementioned application of DESS to operator control of multiple robots, they have successfully been applied to other supervisory control domains such as air traffic control [19]. However, as previously mentioned, the existing models of multiple robot control did not explicitly include operator cognitive inefficiencies either as an input or an output.

A primary limiting factor in single operator-multiple UV systems is operator workload. Indeed, this limitation on the control of multiple UVs extends to any supervisory control system that requires divided attention across multiple tasks such as air traffic control and even supervisors multitasking in a command center such as an air operation center.

Mental workload results from the demands that a task imposes on the operator’s limited resources; it is fundamentally determined by the relationship between resource supply and task demand [14]. Although a number of different ways to measure workload are available [20], [21], given the temporal nature of supervisory control systems, particularly those in multi-UV control, we use utilization as a proxy for measuring mental workload. Utilization is a term found in systems engineering settings and refers to the percentage of busy time of an operator, *i.e.*, given a time period, the percentage of time that a person is busy. In supervisory control settings, it is generally meant as the time that an operator is directed by external events to complete a task (*e.g.*, replanning the path of a UV because of an emergent target). Compared to more common measures of workload [*e.g.*, pupil dilation and the National Aeronautics and Space Administration Task Load Index (NASA TLX)], utilization provides the means of nonintrusively assessing workload in real time. 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. Although, arguably, utilization is not a perfect measure of mental workload, another strength of such a measure is its ratio scale, which allows it to be used in quantitative models.

The concept of utilization as a mental workload measure has been used in numerous studies that examine supervisory controller performance [19], [22], [23]. These studies generally support that, when tasked beyond 70% utilization, operators’ performances decline. In terms of operator attention, high levels of arousal have been shown to induce perceptual narrowing [24].

Although empirically not well established, there is reason to anticipate a decrease in performance with low and high levels of utilization. Previous research suggests an inverted-U shape between arousal/workload level and performance [25]–[27], indicating a decrease in performance with both low and high levels of arousal, which can occur as a function of utilization. With regard to attention, it has been established that vigilance decrement occurs when low arousal is experienced for extended periods of time [28].

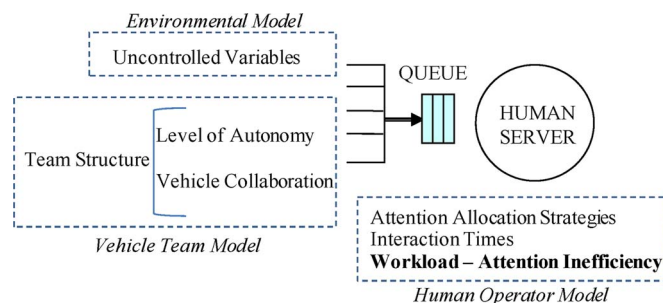


Fig. 1. High-level representation of the discrete event simulation model.

Given the previous research that shows that supervisory control performance drops when utilization is greater than 70% and that there might be performance declines at high and low levels of utilization, we investigated whether the relationship between utilization and performance can be used to not only describe observed human behavior but also to predict it. However, rather than directly connecting workload to performance, we captured effects of workload through delays introduced to the system by humans, which is more appropriate for incorporation in a DES model of human–UV systems.

Previous queuing theory-based human information processing models have also used server utilization as a way to model workload [29], [30]. Although these models have successfully predicted workload and performance, the level of information processing detail captured was at the perceptual level, and the human was represented by multiple servers. Supervisory control of complex systems requires operators to handle high-level tasking through reasoning and judgment, and these tasks are better modeled at a higher (more abstract/cognitive) level, as will be discussed later. Assessing the relation between workload and performance provides a parsimonious way to incorporate workload effects in high-level models of human system interaction when detailed level information processing models are not available.

To address the general explicit lack of accounting for human cognitive inefficiencies in models of human–UV interactions, this paper presents a queuing theory-based DES model of a single operator supervising multiple heterogeneous UVs that includes a utilization-attention inefficiency component.

III. DES MODEL OF UV SUPERVISORY CONTROL

The proposed model utilizes the queuing theory to build a DES model by capitalizing on the event-driven nature of human–supervisory control (see Fig. 1). This section presents an overview of the DES model and the details relevant to the focus of this paper. Further details can be found in [31].

The human operator, responsible for multiple UV supervision, is modeled as a server in a queuing system, with discrete events representing both endogenous and exogenous situations that an operator must address. Endogenous events, which are vehicle generated or operator induced, are events internally created within the supervisory control system, such as when an operator elects to replan an existing UV path to reach a goal in a shorter time. Note that this interaction may not be required by the system and can therefore be operator induced. Events that result from unexpected external environmental conditions

that create the need for operator interaction are defined as exogenous events, such as emerging threat areas that require replanning vehicle trajectories.

The design variables that serve as inputs to the model in Fig. 1 are composed of variables related to the vehicle team (team structure, level of autonomy, and vehicle collaboration), the human operator (ITs, operator attention allocation strategies, and operator workload/attention inefficiency), and a model of environment unpredictability. These variables are discussed in more detail as follows.

A. Vehicle Team Input Variables

Team structure represents the number and type of vehicles included in the system. By representing each vehicle through a distinct input stream, the model can capture heterogeneous team composition, because it includes different arrival processes for events associated with different vehicles. This approach is similar to the previously discussed concept of NT, except that in this case, the NT is for a specific event and not for the whole vehicle (i.e., other events associated with the same vehicle can still be generated, whereas a specific event type is being neglected). Because NTs represent the time that a vehicle can operate without human intervention, they effectively represent degrees of autonomy, i.e., the longer the NTs, the more autonomous the vehicle. Last, the model captures the effect of vehicle collaboration by taking into account the effect of servicing a particular event that belongs to one vehicle on the arrival process of another event that belongs to another vehicle. The types of vehicle autonomy and vehicle collaboration are influenced by the team structure, i.e., the different vehicle types.

B. Human Operator Input Variables

Our DES model represents the human server at a high level, holistically and stochastically capturing human performance through event service times (measured as the time from when operators engage a task to when they finish it), attention allocation strategies (i.e., strategies in choosing what task to service next), and attention inefficiencies.

The length of time that it takes the operator to deal with an event, also known as IT, is captured through a distribution of event service times. ITs occur for a single vehicle task; therefore, to model the effect of an operator controlling multiple vehicles, the model should consider how and when operators elect to attend to the vehicles, also known as attention allocation [32].

The model in Fig. 1 captures two attention allocation strategies that can impact the effectiveness of human–UV interaction. The first strategy is the amount of operator-initiated replanning. Because this model supports endogenous events that are both vehicle generated and operator induced, the rate at which operator-induced events arrive to the system depends on the operator's desire to interact with the vehicles beyond unavoidable vehicle-generated events.

Second, the queuing policy that is used by the operator, i.e., which task waiting in a queue the operator elects to service, is also represented. Examples include the first-in–first-out (FIFO)

queuing scheme and the highest attribute first (HAF) strategy [33]. The HAF strategy is similar to a preemptive priority scheme in that high priority events are serviced first, except that there is no preemption. Therefore, if an event is generated with a priority higher than any of the events already in the system, it will be moved to the front of the queue but will not preempt a lower priority event that is already being serviced.

The need for the last human operator input, the workload-attention inefficiency component, is the hypothesis of this paper and will be discussed in more detail in a subsequent section.

C. Arrival Events

As previously discussed, there are two categories of arrival events: 1) endogenous and 2) exogenous. Endogenous events are events created either by a vehicle (e.g., a UV requests permission from an operator to move to the next target) or by a human (a human initiates a new route without prompting by the system). Because an endogenous event associated with a specific vehicle generally requires attention before an event of the same type can be generated, the arrival process is one of correlated arrivals. For example, if an operator elects to initiate a path replan for vehicle A, he/she must finish servicing that event before again electing to replan vehicle A's path. To capture this phenomenon, the model uses a closed queuing network paradigm such that each endogenous event type in the system (where each endogenous event type is associated with a specific vehicle) has a population of one [34]. The arrival process can therefore be described by a probabilistic distribution over a random variable, which represents the time between the completion of a service for an event associated with a specific vehicle and the next event that the vehicle generates.

Exogenous events stem from sources external to the vehicle (e.g., weather and enemy movements). For example, many emergent threats can simultaneously arise, each requiring operator intervention, thus creating a queue. Therefore, the arrival process in the case of exogenous events is generally one of independent arrivals. The arrival process can therefore be described by a probabilistic distribution over a random variable that represents the interarrival times between exogenous events.

To capture interaction between different event types, the servicing of one event type can be modeled to have an effect on the arrival process of another. For example, a UV might be modeled through two event types: 1) the need for operator interaction whenever the operator is required to identify a target and 2) the need for operator interaction once target identification is complete and the vehicle requires a new assignment. In this case, event type 2 is generated only after event type 1 is serviced by the operator.

By modeling the operator as a single server, this model assumes serial operator interaction such that events that arrive while the operator is busy will wait in a queue. Although it is possible for humans to multitask, the appropriateness of the human model depends on the level of processing detail under consideration. When considering supervisory control tasks for complex systems such as those in UV systems, humans are generally required to handle high-level tasking that involves application of human judgment and reasoning. Although

operators can rapidly switch between cognitive tasks, any sequence of tasks requiring complex cognition will form a queue, and consequently, wait times will build [13]. As such, humans will act as serial processors in that they can only solve a single complex task at a time [35], [36]. The serial processing model has been applied to capture higher level tasking [19], whereas the parallel processing has been more generally applied for capturing lower level perceptual processing such as that involved in driving tasks [18], [37]. For example, Schmidt [19] has suggested a single-server queuing system as appropriate for modeling an air traffic controller in charge of conflict assessment and resolution, as well as general routing type tasks.

Based on the assumption of serial operator interaction, the service processes can be described by probabilistic distributions representing the interval from the time the operator begins to service a particular event until the operator is done servicing (this approach applies to both endogenous and exogenous events).

IV. EXPERIMENTAL MOTIVATION FOR A DES WORKLOAD MODEL

An online experiment was conducted to validate the previously described DES model for different vehicle team structures, with the ultimate goal of predicting performance of various human–UV systems. Prior to this experiment, there was no workload-attention inefficiency component in the DES model. Moreover, the initial model validation in [31] was conducted on experimental data from 16 participants. Given the relatively small sample size of the previous validation experiment, we wanted to ensure that any significant trends that result from the current experiment had higher statistical power. Thus, online data collection was chosen as a means of increasing our sample size. However, to decrease the likelihood of participant withdrawal from the online experiment, we kept the trials fairly short at 10 min. To ensure the validity of online experimentation, a pilot study was conducted with 15 participants completing the experiment online and an additional 15 completing it in a laboratory setting with an experimenter present. No significant differences were observed between the two groups for the variables of interest [38].

As will be shown through the experimental results, the model without considering a workload-attention inefficiency relationship does not adequately replicate results of the human-in-the-loop trials. Model inaccuracies provided the motivation to incorporate the human delays in the DES model.

A. Participants

A total of 74 participants, six females and 68 males, aged 18–50 completed the study. There were 36 participants between the ages of 18 and 25, 30 participants between 26 and 35, and eight participants whose age was greater than 35. The majority of participants were students, and some were UV researchers from the industry ($n = 14$). Participants were randomly assigned to the experimental conditions based on the order that they logged in to the online server. The breakdown of UV researchers across different experimental conditions was fairly constant ($n = 4, 5, 5$). There was no monetary compensation

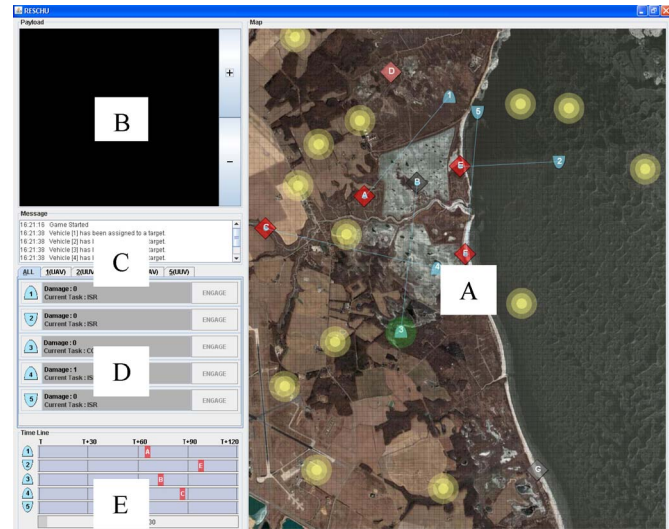


Fig. 2. RESCHU interface. (a) Map. (b) Camera window. (c) Message box. (d) Control panel. (e) Timeline.

for participation; however, the best performer received a \$200 gift certificate.

B. Experimental Test Bed

The research environment for supervisory control of heterogeneous UVs (RESCHU) simulator was used in the experiment and allowed operators to control a team of UVs composed of unmanned air vehicles (UAVs) and unmanned underwater vehicles (UUVs). All vehicles were engaged in surveillance tasks, with the ultimate mission of locating specific objects of interest in urban coastal and inland settings. Although there was only a single UUV type, there were two UAV types: 1) a UAV type that provided high-level sensor coverage (akin to a Global Hawk UAV) and 2) a UAV type that provided more low-level target surveillance and video gathering (similar to a Predator UAV). Thus, there were three different vehicle types under control for a single operator. Because previous research has shown that, to allow for the simultaneous supervision and payload management (e.g., managing cameras for target identification) of multiple UVs, navigation tasks for the different vehicles should be highly automated [6], this condition was a basic assumption for this simulation.

The RESCHU interface consisted of five major sections (see Fig. 2). The map displayed the locations of vehicles, threat areas, and areas of interests (AOIs) [see Fig. 3(a)]. Vehicle control was 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) were displayed in the message box, along with a timestamp [see Fig. 3(b)]. When the vehicles reached an AOI, a simulated video feed was displayed in the camera window. The participant had to visually identify a target in this simulated video feed. Example targets and objects of interest included cars, swimming pools, and helipads.

The control panel provided vehicle health information and information on the vehicle’s mission [see Fig. 3(c)]. The

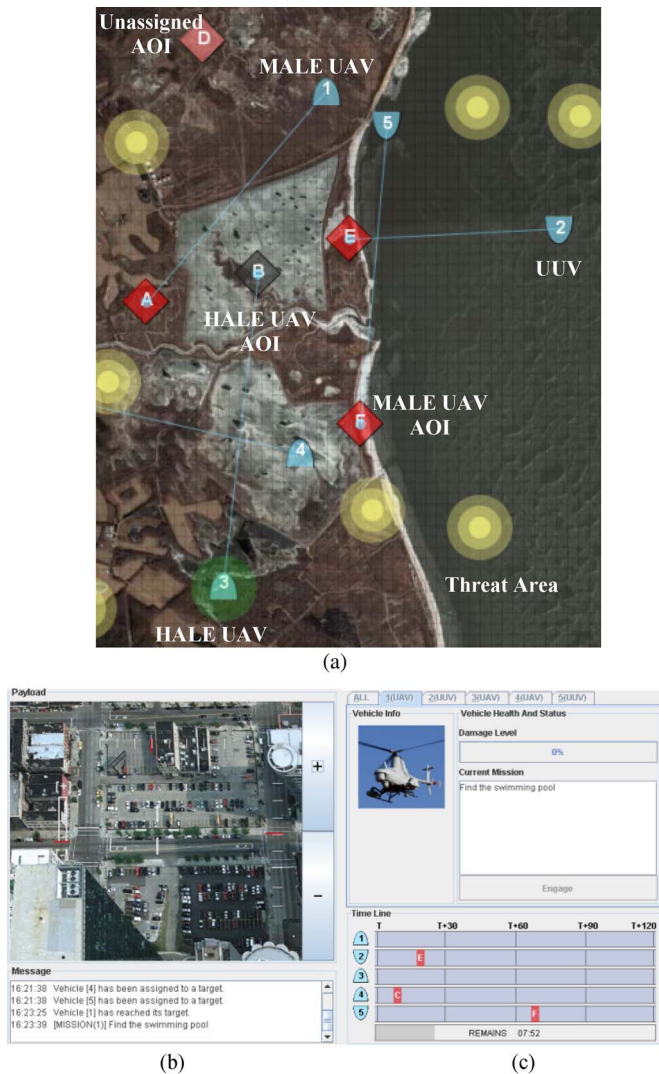


Fig. 3. (a) Section of the map. (b) Activated camera view and message box (c) Control panel and timeline.

timeline displayed the estimated time of arrival to waypoints and AOIs. Beneath the timeline was a mission progress bar that showed the amount of time remaining in the total simulation.

As previously discussed, three types of vehicles were used in this experiment: 1) a high-altitude long-endurance (HALE) UAV; 2) medium-altitude long-endurance (MALE) UAVs; and 3) UUVs. Both the MALE UAVs and the UUVs traveled to the AOIs [red AOIs in Fig. 3(a)] with a predetermined target that needed to be visually acquired by the operator. MALE UAVs can travel to any AOI (both shore and land), whereas UUVs can only travel to AOIs on the shoreline. A HALE UAV traveled to AOIs that did not yet have a target specified [gray AOIs in Fig. 3(a)] and carried a synthetic aperture radar (SAR)-type sensor, which allowed for target specification. These newly discovered targets were later acquired by a MALE UAV or a UUV.

When the vehicles completed their assigned tasking, an automated-path planner automatically assigned the HALE UAV to an AOI that needed intelligence and the MALE UAVs and UUVs to AOIs with predetermined targets. The automatically assigned AOIs were not necessarily the optimal choice. The operator could change the assigned AOI and avoid threat areas

by changing a vehicle's goal or adding a waypoint to the path of the vehicle to go around the threat area.

When a vehicle arrived to an AOI, a visual flashing alert indicated that the operator could engage the payload. For a HALE UAV, clicking the engage button resulted in the uncovering of the target in the AOI. For a MALE UAV or a UUV, engaging the payload caused the camera window to display the simulated live video feed [see Fig. 3(b)]. The operator then had to complete a search task by panning and zooming the camera until the specified target was located. After the operator had submitted the target identification, the message box notified the operator on the accuracy of response (used to simulate feedback that real operators get from their commanders or teammates as a consequence of their actions), and the vehicle was automatically reassigned to a new AOI.

Participants were instructed to maximize their score in the following four ways: 1) avoid threat areas that dynamically changed; 2) complete as many of the search tasks correctly; 3) take advantage of replanning, when possible, to minimize vehicle travel times between AOIs; 4) ensure that a vehicle was always assigned to an AOI, whenever possible.

The UVs were not modeled on real UV performance data, because this experiment simulated a futuristic system, i.e., there are no operational command and control systems with integrated heterogeneous unmanned operations. However, to create realism, UUVs were modeled to move slower than UAVs based on typical current platform capabilities. In the experiment, the UVs required human intervention multiple times, thus creating a fast-paced scenario and represented high-workload situations.

C. Experimental Design and Independent Variables

The experiment was a completely randomized design with vehicle team heterogeneity level as a between-subject factor: none ($n = 26$), medium ($n = 25$), and high ($n = 23$). The no-heterogeneity condition included five MALE UAVs. The medium-heterogeneity condition had three MALE UAVs and two UUVs. Because the UUVs were slower than UAVs, they produced events less frequently. The maximum level of heterogeneity required managing two MALE UAVs, two UUVs, and one HALE UAV. HALE UAVs were restricted to gray AOIs, which appeared at a ratio of five to two, compared to red AOIs, which the UUVs and MALE UAVs could visit without assistance from the HALE. Thus, the arrival rates of events for HALE UAVs were different from both the MALE UAVs and UUVs. Moreover, service times were different, because the HALE UAVs required just milliseconds of service time (operators clicking the engage button). Last, 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, followed by the medium- and then the high-heterogeneity condition.

D. Procedure

The online experiment began with an interactive tutorial, followed by an open-ended practice session. The interactive

tutorial had to be completed before the participants could start the practice session. During the interactive tutorial, the participants had to repeat a task until they correctly performed it. Thus, a major part of the training took place during the interactive tutorial. After participants felt comfortable with the task and the interface, they could end the practice session and start the 10-min experimental session. Pilot participants were observed to spend, on the average, 10 min doing the practice session. The website was password protected, and participation was only through invitation. All data were recorded to an online database. Demographic information was collected through a questionnaire presented before the tutorial.

V. DES REPLICATIONS WITHOUT MODELING WORKLOAD EFFECTS

This section presents the results obtained from the experiment, followed by the analysis of the model’s ability to describe the observed data and predict how changes in the vehicle heterogeneity structure will alter variables of interest.

The variables of interest for evaluating model predictions were score, average search task wait time, and UT. Mission performance was assessed through score, which 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. Search task wait time was calculated from vehicle arrival to an AOI to the operator engaging the search task. Average search task wait time assessed system performance efficiency, because it demonstrated the effects of operator inefficiencies through system delays. UT was calculated as the proportion of time that the operator actively interacted with the display (e.g., adding a way point and engaging in a visual task) during the course of the experiment. Utilization therefore excluded any monitoring time expended by operators.

A. Observed Effects

A preliminary analysis demonstrated significant correlations between the three variables of interest: 1) utilization/score ($\rho = -.25, p = .03$); 2) utilization/search task wait times ($\rho = .50, p < .0001$); and 3) score/search task wait times ($\rho = -.58, p < .0001$). Because these three measures are correlated, multivariate analysis of variance (MANOVA) was performed to control for the inflation of Type I error. Significant findings were followed with univariate analysis to assess the magnitude of the effect that vehicle heterogeneity level had on each variable.

The MANOVA results indicated that there were significant effects of heterogeneity level (Wilks’ Lambda = 0.4, $F(6, 138) = 13.33, p < .0001$). The univariate analysis suggested that the effect of heterogeneity level is attributable to the differences observed in all of the three variables of interest (see Fig. 4). There were significant differences in utilization between different heterogeneity levels ($F(2, 71) = 33.31, p < .0001$), score ($F(2, 71) = 8.13, p = .0007$), and search task wait times ($F(2, 71) = 7.75, p = .0009$).

The pair-wise comparisons (see Table I) revealed that, as expected due to the tempo of arriving events, utilization was the highest for the no-heterogeneity level, followed by the medium- and high-heterogeneity levels. UAVs, which spent a

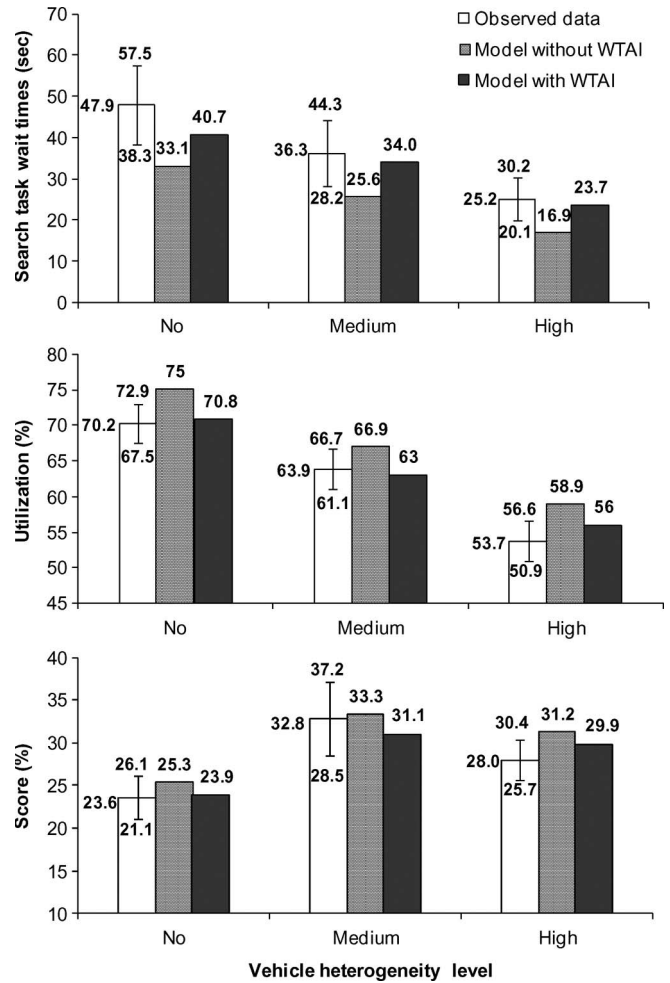


Fig. 4. Observed data with 95% confidence intervals and model replications for the three dependent variables of interest. (a) Search task wait times. (b) Utilization. (c) Score.

TABLE I
PAIRWISE COMPARISONS FOR DEPENDENT MEASURES

Comparison of heterogeneity levels	Difference	t-value (df: 71)	p	95% CI
<i>Search task wait times</i>				
No vs. medium	11.63 s	2.06	.04	(0.35, 22.92)
No vs. high	22.74 s	3.93	.0002	(11.21, 34.27)
Medium vs. high	11.11 s	1.90	.06	(-0.53, 22.75)
<i>Utilization</i>				
No vs. medium	6.29 %	3.17	.002	(2.34, 10.25)
No vs. high	16.45 %	8.12	<.0001	(12.41, 20.49)
Medium vs. high	10.16 %	4.97	<.0001	(6.08, 14.24)
<i>Score</i>				
No vs. medium	-9.24 %	-4.03	<.0001	(-13.8, -4.67)
No vs. high	-4.42 %	-1.89	.06	(-9.08, 0.25)
Medium vs. high	4.82 %	2.04	.045	(0.11, 9.53)

considerable amount of time underwater, required less frequent interaction with the human operator than UAVs. In addition, HALE UAVs required shorter interactions than the MALE UAVs. Thus, as the level of heterogeneity increased, operators interacted less frequently with the vehicles due to longer NTs and shorter interaction periods. This cascading effect was also shown in the wait time metric, because the homogeneous (no heterogeneity) team structure generated significantly longer

TABLE II
EVENT ARRIVAL AND SERVICE DISTRIBUTIONS FOR THE THREE HETEROGENEITY VEHICLE STRUCTURES

Vehicle	Event type	Event generator	Distribution (parameters)			
			Exp: Exponential, Γ : Gamma, Log-N: Log-normal, N: Normal, NA: Not Applicable			
			High Heterogeneity	Medium Heterogeneity	No Heterogeneity	
MALE UAV	Type1	Search task arrival	Γ (α : 4.61, β : 21.97)	Γ (α : 4.04, β : 26.80)	Γ (α : 4.08, β : 26.27)	
		Modified search task arrival due to re-plan	Log-N (μ : 3.86, σ : 0.54)	Γ (α : 3.21, β : 19.67)	Γ (α : 2.80, β : 18.65)	
		Search task service	Log-N (μ : 3.14, σ : 0.59)	Log-N (μ : 2.96, σ : 0.64)	Log-N (μ : 2.94, σ : 0.63)	
	Type2	Re-plan ratio	Bernoulli (p: .58)	Bernoulli (p: .51)	Bernoulli (p: 0.48)	
		Re-plan service	N (μ : 3.19, σ : 7.32)	Exp (λ : 2.52)	Exp (λ : 3.2)	
	Type3	Idle ratio	Bernoulli (p: .1)	Bernoulli (p: 0)	Bernoulli (p: 0)	
		Idle duration	Exp (λ : 34.88)	NA	NA	
		Idle service	N (μ : 3.19, σ : 7.32)	NA	NA	
	Type4	Threat area arrival	Exp (λ : 105.49)	Exp (λ : 95.15)	Exp (λ : 72.9)	
		Threat area service	Log-N (μ : 0.75, σ : 0.56)	Γ (α : 1.64, β : 1.49)	Γ (α : 0.58, β : 0.56)	
	UUV	Type1	Search task arrival	Γ (α : 2.89, β : 73.51)	N (μ : 174.07, σ : 102.33)	
			Modified search task arrival due to re-plan	Log-N (μ : 4.73, σ : 0.58)	Log-N (μ : 4.61, σ : 0.37)	
Search task service			Log-N (μ : 2.95, σ : 0.69)	Log-N (μ : 2.78, σ : 0.66)		
Type2		Re-plan ratio	Bernoulli (p: 1)	Bernoulli (p: 1)		
		Re-plan service	N (μ : 3.19, σ : 7.32)	Exp (λ : 2.52)		
Type3		Idle ratio	Bernoulli (p: .69)	Bernoulli (p: .42)		
		Idle duration	Exp (λ : 35.91)	Exp (λ : 59.1)		
		Idle service	N (μ : 3.19, σ : 7.32)	Exp (λ : 2.52)		
Type4		Threat area arrival	Exp (λ : 182.8)	Exp (λ : 168.63)		
		Threat area service	Log-N (μ : 0.75, σ : 0.56)	Γ (α : 1.64, β : 1.49)		
HALE UAV		Type1	Search task arrival	N (μ : 154.38, σ : 56.05)		
			Modified search task arrival due to re-plan	N (μ : 94.42, σ : 39.57)		
	Search task service		N (μ : 0.1, σ : 0.1)			
	Type2	Re-plan ratio	Bernoulli (p: .38)			
		Re-plan service	N (μ : 3.19, σ : 7.32)			
	Type4	Threat area arrival	Exp (λ : 151)			
		Threat area service	Log-N (μ : 0.75, σ : 0.56)			

search task wait times compared to both the medium- and the high-heterogeneity team structures. Because the operators had to interact more often with the MALEs, the no-heterogeneity service queues were larger, thus generating longer wait times.

Although the wait time and utilization results were expected due to the decrease in ITs and longer NTs with increasing heterogeneity, the performance score results in Fig. 4 showed a different trend. The medium-heterogeneity configuration resulted in significantly higher score than both the no- and high-heterogeneity team structures. These results suggest that, although increasing the variety of the vehicles under control with different capabilities can reduce operator workload and system delays if NTs are increased, the use of higher levels of autonomy can also lead to degraded performance. Although this finding is important and, no doubt, has significant implications, we leave this result as an area of future research. The focus of this paper is to develop a DES model that can both replicate these results and predict the likely outcomes of other team configurations.

B. DES Replications Without Workload Effects

Using the participant data from the experiment, DES models were constructed for the three vehicle heterogeneity conditions.

In RESCHU there were four different vehicle event types that required user interaction: 1) a vehicle arrives to an AOI and requires the operator to undertake a search task (a vehicle-generated endogenous event); 2) an opportunity for replanning the vehicle's path to a closer AOI (an operator-induced endogenous event); 3) an idle vehicle that requires assignment to an AOI (a vehicle-generated endogenous event); and 4) the intersection of a vehicle's path with a threat area (an exogenous environmental event). Table II presents the fitted distribution types and their parameters for these four different event arrivals and services. All distributions were generated from experimental data using distribution fitting software, assessed with the Kolmogorov–Smirnov goodness-of-fit tests. Using these distributions, 10 000 trials were conducted for each DES model.

The probabilistic distribution parameters in Table II constitute the complete list of parameters used in the DES model. When replicating the high-heterogeneity team structure, the model used the parameter estimates from the high-heterogeneity experimental data (column 4 in Table II). Similarly, the medium- and no-heterogeneity conditions were replicated using the parameters obtained from the medium-heterogeneity (column 5) and no-heterogeneity (column 6) data, respectively.

As shown in Fig. 4, the model estimates for the three vehicle heterogeneity levels do not fall in the 95% confidence intervals obtained from the observed data for both the search task wait times and UT. Given that there were 10 000 trials run for the DES model, the standard errors for the estimated model means were practically 0. Thus, when the estimated means fall outside the 95% confidence intervals of the observed means, there is a statistically significant difference between the two. Therefore, this model fails to accurately replicate the observed data, which means that the model cannot also accurately predict for other UV team combinations. The following sections attempt to improve model replication by incorporating the workload effects on operator attention inefficiencies.

VI. UTILIZATION AND ATTENTION INEFFICIENCIES

As previously discussed, UT, our measure of workload, is hypothesized to affect performance such that it is degraded at both high and low ends of the utilization curve. Utilization can guide how well operators notice events, inducing unnecessary wait times for vehicle servicing, in particular through attention switching delays (i.e., WTAI).

In the experiment, WTAI was measured as the time from an emerging threat area intersection with a vehicle's path to the time when the participant responded to this intersection. The response to emerging threat areas was chosen as the measure of WTAI, because avoiding threat areas was the highest priority task for the participants, and it required decisive and identifiable actions.

The experimental data revealed that the overall average post hoc utilization values for the different vehicle heterogeneity levels ranged between 40% and 80%. However, this static post hoc calculation does not reflect the dynamic nature of utilization; therefore, four utilization values were calculated for 2.5-min time windows for the 10-min experiment. The average WTAI for different values of UT across the four time intervals are presented in Fig. 5(a). Due to missing data, the number and spread of utilization bins for each condition differed. In the case of the no-heterogeneity condition, only four bins had enough samples, all at higher utilization values due to the high operational tempo. Fig. 5(b) demonstrates the associated performance scores for these same utilization bins.

To determine if significant differences existed in WTAI across 10% utilization bins for the three different heterogeneity conditions, a repeated-measures analysis of variance (ANOVA) was conducted. A logarithmic transformation was performed on WTAI to meet statistical modeling assumptions, i.e., the normality and homogeneity of variances. Results revealed that there were significant differences between different utilization intervals for the medium-heterogeneity condition ($F(5, 119) = 2.43, p = .04$) and high-heterogeneity condition ($F(5, 121) = 8.11, p < .0001$). Differences in WTAI for the no-heterogeneity condition was only marginally significant ($F(3, 110) = 2.54, p = .06$).

In the no-heterogeneity case, pairwise comparisons showed that 60%–70% utilization resulted in shorter WTAI than both the 80%–90% ($p = .03$) and 70%–80% utilization bins ($p = .02$). In the medium-heterogeneity case, the 80%–90% utiliza-

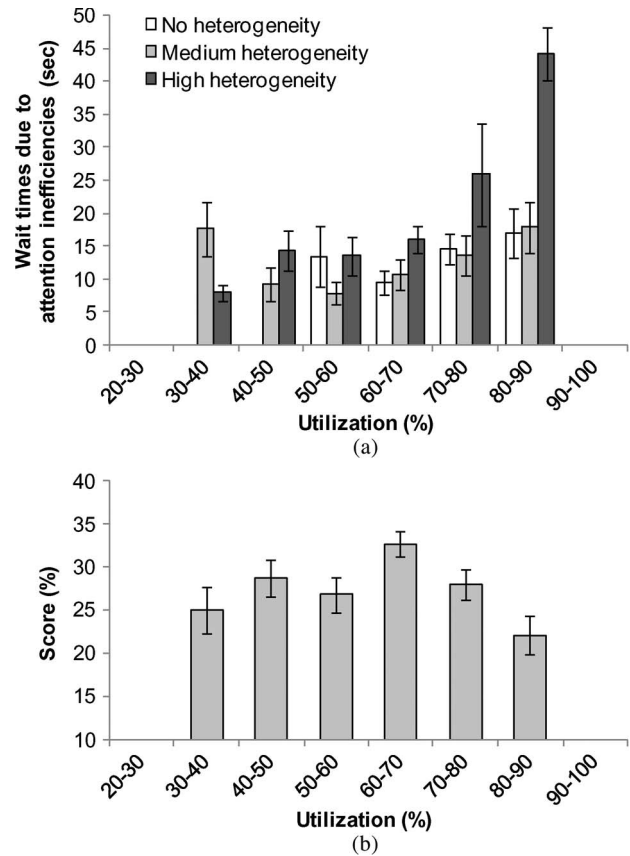


Fig. 5. (a) Experimental results for UT–WTAI relation (with standard error bars). (b) Experimental results for score versus UT (with standard error bars).

tion bin resulted in longer WTAI than for the 60%–70% bin ($p = .03$), 50%–60% bin ($p = .008$), and 40%–50% bin ($p = .04$). In addition, the 30%–40% utilization bin also resulted in longer WTAI than the 60%–70% bin ($p = .047$) and the 50%–60% bin ($p = .02$). Finally, in the high-heterogeneity case, 80%–90% utilization bin resulted in the longest WTAI compared to all other utilization values (70%–80%: $p = .04$; 60%–70%: $p = .001$; 50%–60%: $p < .0001$; 40%–50%: $p < .0001$; 30%–40%: $p < .0001$). However, the 30%–40% utilization resulted in significantly shorter WTAI compared to the 60%–70% ($p = .002$) and 70%–80% utilization ($p = .005$).

These results demonstrate that WTAI is longer at higher utilization levels than at medium utilization levels, which is consistent with our initial hypothesis. However, the medium- and high-heterogeneity conditions contradict in terms of how utilization is related to WTAI for low-utilization levels. Although the medium condition is in agreement with the initial hypothesis that lower utilization values have higher WTAI than medium utilization values, the high-heterogeneity case resulted in the reverse trend. Most of the data points for all three conditions fell toward higher utilization values, and the 10-min experiment did not require vigilance to be maintained for a long period of time; therefore, the functional form of WTAI at lower utilization values is left for future work. Moreover, at high utilization values, WTAI appears to be much higher for the high-heterogeneity vehicle structure. This result suggests that high UT that results from controlling multiple vehicles with different

capabilities can particularly be detrimental to WTAI and, thus, the overall mission performance.

To assess if the UT/WTAI relation was consistent across the four different time windows, another repeated-measures ANOVA was conducted using the experimental data from all three vehicle heterogeneity levels. Vehicle heterogeneity could not be included as a factor in the model due to the large number of missing design cell combinations and the small number of observations in some of the design cells. A logarithmic transformation was performed on WTAI to stabilize variance. The time window ($p = .26$) and the time window–UT interaction were not significant ($p = .71$), suggesting that UT–WTAI relation can be considered as fairly consistent across the different time windows.

The distribution of the performance scores [see Fig. 5(b)] suggests that over-utilization or underutilization caused the overall mission performance to degrade. A mixed linear regression model demonstrated that utilization was significantly associated with score ($F(5, 126) = 4.2, p = .001$), given a backward selection model, controlling for vehicle heterogeneity level ($F(2, 71) = 8.37, p = .0005$), time window ($F(3, 208) = 24.30, p < .0001$), and vehicle heterogeneity level–time window interaction ($F(6, 208) = 2.08, p = .06$).

Pairwise comparisons for Fig. 5(b) revealed that 60%–70% utilization corresponded to significantly higher scores than most other utilization values (30%–40%: $p = .02$; 50%–60%: $p = .02$; 70%–80%: $p = .02$; 80%–90%: $p < .0001$). In addition, 80%–90% utilization resulted in lower 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, performance significantly degrades [19], [22]; therefore, these results also support that there is a threshold for performance in terms of UT.

VII. DES REPLICATIONS WITH MODELING WORKLOAD EFFECTS

The aforementioned experimental data were used to incorporate the effects of workload in the DES model. This revision of the DES model included modification of the arrival of events (both exogenous and endogenous) so that events arrive to the system after they are noticed by the operator. Thus, an operator's inattention efficiency due to workload was stochastically modeled by introducing vehicle wait times as a function of utilization. Whenever there was an event arrival, the utilization for the previous 2.5-min time window was calculated, which, in turn, was used to identify the appropriate time penalty from the UT–WTAI relation, as presented in Fig. 5(a).

As shown in Fig. 4, the revised DES model estimates for the three dependent variables (i.e., search task wait times, utilization, and score) fall in the 95% confidence intervals obtained from the observed data. Therefore, the revised DES model, which accounts for WTAI, more accurately estimates the observed data for all vehicle heterogeneity levels, particularly compared to the results without including this relationship. This result suggests that WTAI can be inserted into a DES model as a function of UT, which can be fed back through the

model, providing a better estimate of the operator's influence on the system.

VIII. DES PREDICTIONS WITH MODELING WORKLOAD EFFECTS

This section presents the effects of WTAI on the model's predictive power, i.e., the model's ability to predict the effects of changes in the vehicle heterogeneity structure. To assess predictive power, the DES model was constructed using the experimental data for the medium-heterogeneity condition to predict for the no- and high-heterogeneity levels. Therefore, model distributions (e.g., service times and UT–WTAI relation) were populated based on this experimental data subset only. The medium-heterogeneity condition was chosen to build the model, because we wanted to assess the model's predictive capabilities for both increasing and decreasing heterogeneity.

Going from the medium-heterogeneity team to the no-heterogeneity team, the change required replacement of two UUVs by two MALE UAVs. Thus, in the DES, the appropriate arrival and service processes were substituted. The parameter estimates for MALE UAVs from column 5 in Table II were used to predict for the no-heterogeneity vehicle team structure.

Going from the medium-heterogeneity team to the high-heterogeneity team, one of the MALE UAVs was replaced by a HALE UAV. The parameter estimates for MALE UAVs and UUVs from column 5 in Table II were used when predicting for the high-heterogeneity team structure, in particular when modeling MALE UAVs and UUVs. However, simple arrival distribution substitution was not possible for HALE UAVs, because there were no HALE UAVs in the medium-heterogeneity team, and therefore, the arrival processes of vehicle-generated events could not be derived for this type of vehicle. A Monte Carlo simulation was used to derive the missing data. More specifically, Monte Carlo simulations were used to derive travel times between randomly located AOIs, which translated to a new vehicle-generated event arrival. The samples from the simulations were then used to build the arrival distributions for HALE UAVs. The replan ratio for HALE UAVs was also obtained from Monte Carlo simulation by identifying replan opportunities based on the availability of closer targets. The service times for HALE UAVs were assumed to be negligible, because servicing HALE UAVs only required clicking on a button, which took much shorter than servicing MALE UAVs or UUVs.

As previously discussed, the degree of heterogeneity in team structure resulted in significant differences in UT, search task wait times, and score. The DES model that incorporates workload more accurately predicted these observed changes in search task wait times and the UT than did the DES model that did not account for workload (see Fig. 6). In the case of the performance score and utilization, the predictions were accurate for the homogeneous condition, but for the high-heterogeneity condition, the revised model's estimates were not as accurate. This inaccuracy is likely due to the missing data problem. However, the DES model still captured the trend of increasing performance score as the team structure changed from no heterogeneity to high.

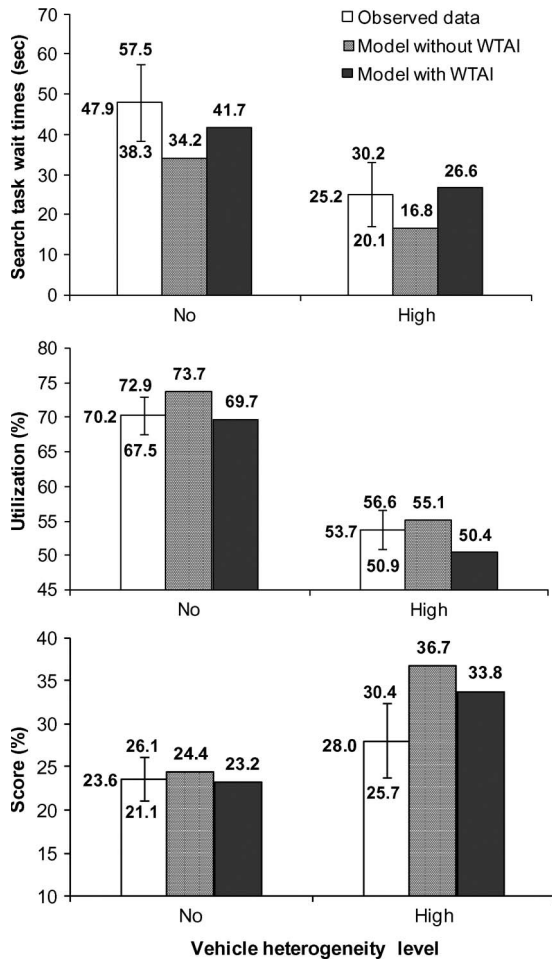


Fig. 6. Prediction: Observed data with 95% confidence intervals and model predictions for the three dependent variables of interest: (a) search task wait times, (b) utilization, and (c) score.

IX. CONCLUSION

This paper has demonstrated the incorporation of the effects of workload in a DES model of human supervisory control, in particular multiple UV supervisory control. As demonstrated in a human-in-the-loop experiment, system delays caused by operator attention inefficiencies (measured through attention switching delays) are significantly related to UT, and these system delays can negatively impact the overall mission. High levels of workload (measured through utilization), and, in some cases, low workload led to increased attention switching delays. Consequently, system wait times increased, which ultimately led to poor mission performance. The DES model with the inclusion of these additional wait times due to operator attention inefficiencies provided enhanced accuracy for replicating experimental observations and predicting results for different UV team structures compared to the DES model without accounting for operator workload.

Note that this DES model does not capture all possible human cognitive inefficiencies but rather an aggregate effect of inefficient information processing, which likely has many sources that exist at a level difficult to capture in a DES model. There are likely many more sources of cognitive inefficiencies, such as operator trust [39], that are manifested through

system delays. However, in the spirit of Occam's razor, we have focused on a single quantifiable relationship that provides bounded estimates of operator behavior.

One issue that requires further investigation is the temporal and dynamic nature of utilization. All utilization metrics reported here were post hoc aggregate measures, which are not accurate as an instantaneous measure of workload. Thus, more research is needed to determine how utilization can be measured in a real-time 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 utilization. The effects of low utilization or cognitive underload and the nature of sustained and variable rates of utilization changes also deserve further scrutiny.

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

The methodology used in this paper presents a way of dynamically adjusting performance scores based on operator workload to make generally effective system predictions through DES models. The DES model was substantially improved considering the effects of workload; therefore, this paper has implications toward developing more realistic models of human supervisory control and human-system performance.

ACKNOWLEDGMENT

The authors would like to thank Yale Song for his help in coding the experimental test bed and Paul de Jong for his help in data processing.

REFERENCES

- [1] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press, 1992.
- [2] DOD, "Unmanned Systems Roadmap," Office of the Secretary Defense, Washington, DC, 2007.
- [3] D. S. Alberts, J. J. Garstka, and F. P. Stein, *Network Centric Warfare: Developing and Leveraging Information Superiority*, 2nd ed. Washington, DC: CCRP Publication Series, 1999.
- [4] A. Feickert, The Army's Future Combat System (FCS): Background Issues for Congress, Congressional Research Service RL32888, 2008.
- [5] M. L. Cummings, C. E. Nehme, J. W. Crandall, and P. J. Mitchell, "Predicting operator capacity for supervisory control of multiple UAVs," in *Innovations in Intelligent Machines*, J. S. Chahl, L. C. Jain, A. Mizutani, and M. Sato-Ilic, Eds. New York: Springer-Verlag, 2007, pp. 11–36.
- [6] M. L. Cummings, S. Bruni, S. Mercier, and P. J. Mitchell, "Automation architecture for single operator, multiple UAV command, and control," *Int. Command Control J.*, vol. 1, no. 2, pp. 1–24, 2007.
- [7] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, 3rd ed. Boston, MA: McGraw-Hill, 2000, ser. McGraw-Hill International Series.
- [8] M. Woolridge, *Introduction to Multiagent Systems*, 2nd ed. New York: Wiley, 2009.

- [9] Y. Shoham and K. Leyton-Brown, *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [10] J. W. Crandall, M. A. Goodrich, R. O. Olsen, and C. W. Nielsen, "Validating human-robot interaction schemes in multitasking environments," *IEEE Trans. Syst., Man, Cybern. A: Syst., Humans*, vol. 35, no. 4, pp. 438–449, Jul. 2005.
- [11] S. Mau and J. Dolan, "Scheduling to minimize downtime in human-multirobot supervisory control," in *Proc. Int. Workshop Planning Scheduling Space*, 2006, pp. 268–277.
- [12] R. O. Olsen and M. A. Goodrich, "Metrics for evaluating human-robot interactions," in *Proc. NIST Performance Metrics Intell. Syst. Workshop*, 2003. [CD-ROM].
- [13] M. L. Cummings and P. J. Mitchell, "Predicting controller capacity in supervisory control of multiple UAVs," *IEEE Trans. Syst., Man, Cybern. A: Syst., Humans*, vol. 38, no. 2, pp. 451–460, Mar. 2008.
- [14] J. W. Senders, "The human operator as a monitor and controller of multidegree of freedom systems," *Hum. Factors Electron.*, vol. HFE-5, no. 1, pp. 2–5, Sep. 1964.
- [15] T. B. Sheridan, "On how often the supervisor should sample," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-6, no. 2, pp. 140–145, Apr. 1969.
- [16] J. R. Carbonell, J. L. Ward, and J. W. Senders, "A queuing model of visual sampling experimental validation," *IEEE Trans. Man-Mach. Syst.*, vol. MMS-9, no. 3, pp. 82–87, Sep. 1968.
- [17] Y. Liu, "Queuing network modeling of human performance of concurrent spatial and verbal tasks," *IEEE Trans. Syst., Man, Cybern.*, vol. 27, no. 2, pp. 195–207, Mar. 1997.
- [18] 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 Trans. Comput.-Human Interaction*, vol. 13, no. 1, pp. 37–70, Mar. 2006.
- [19] D. K. Schmidt, "A queuing analysis of the air traffic controller's workload," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 6, pp. 492–498, Jun. 1978.
- [20] C. D. Wickens and J. G. Hollands, *Engineering Psychology and Human Performance*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2000.
- [21] W. W. Wierwille and F. T. Eggemeier, "Recommendations for mental workload measurement in a test and evaluation environment," *Hum. Factors*, vol. 35, no. 2, pp. 263–281, Jun. 1993.
- [22] M. L. Cummings and S. Guerlain, "Developing operator capacity estimates for supervisory control of autonomous vehicles," *Hum. Factors*, vol. 49, no. 1, pp. 1–15, Feb. 2007.
- [23] W. B. Rouse, *Systems Engineering Models of Human-Machine Interaction*. New York: North Holland, 1983.
- [24] D. Kahneman, *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, 1973.
- [25] D. O. Hebb, "Drives and the C.N.S. (Conceptual Nervous System)," *Psychol. Rev.*, vol. 62, no. 4, pp. 243–254, Jul. 1955.
- [26] R. M. Yerkes and J. D. Dodson, "The relation of strength of stimulus to rapidity of habit formation," *J. Compar. Neurol. Psychol.*, vol. 18, no. 5, pp. 459–482, Nov. 1908.
- [27] R. W. Proctor and T. Van Zandt, *Human Factors in Simple and Complex Systems*, 2nd ed. Boca Raton, FL: CRC Press, 2008.
- [28] R. Parasuraman and D. R. Davies, "Decision theory analysis of response latencies in vigilance," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 2, no. 4, pp. 569–582, Nov. 1976.
- [29] C. Wu and Y. Liu, "Queuing network modeling of driver workload and performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 528–537, Sep. 2007.
- [30] C. Wu, O. Tshimhoni, and Y. Liu, "Development of an adaptive workload management system using the queuing network-model human processor (QN-MHP)," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 463–475, Sep. 2008.
- [31] C. Nehme, B. Mekdeci, J. W. Crandall, and M. L. Cummings, "The impact of heterogeneity on operator performance in futuristic unmanned vehicle systems," *Int. C2 J.*, vol. 2, pp. 1–30, Nov. 2008.
- [32] J. W. Crandall and M. L. Cummings, "Identifying predictive metrics for supervisory control of multiple robots," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 942–951, Oct. 2007.
- [33] M. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [34] E. D. Lazowska, J. Zahorjan, G. S. Graham, and K. C. Sevcik, *Quantitative System Performance: Computer System Analysis Using Queuing Network Models*. Upper Saddle River, NJ: Prentice-Hall, 1984.
- [35] D. E. Broadbent, *Perception and Communication*. Oxford, U.K.: Pergamon, 1958.
- [36] A. T. Welford, "The psychological refractory period and the timing of high-speed performance: A review and a theory," *Brit. J. Psychol.*, vol. 43, pp. 2–19, 1952.
- [37] C. Wu, Y. Liu, and C. M. Quinn-Walsh, "Queuing network modeling of a real-time psychophysiological index of mental workload—P300 in event-related potential (ERP)," *IEEE Trans. Syst., Man, Cybern. A: Syst., Humans*, vol. 38, no. 5, pp. 1068–1084, Sep. 2008.
- [38] C. Nehme, "Modeling human supervisory control in heterogeneous unmanned vehicle systems," Ph.D. dissertation, MIT, Cambridge, MA, 2009.
- [39] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004.



Birsan Donmez (M'10) received the B.S. degree in mechanical engineering from Bogazici University, Istanbul, Turkey, in 2001, the M.S. degree in 2004, and Ph.D. degree in industrial engineering and the M.S. degree in statistics from the University of Iowa, Iowa City, both in 2007, respectively.

She is currently an Assistant Professor with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada. Before joining the University of Toronto, she was a Postdoctoral Associate with the Massachusetts Institute of Technology, Cambridge. Her research interests are centered on understanding and improving human behavior and performance in multitask and complex situations using a wide range of analytical techniques.



Carl Nehme received the Bachelor's degree in computer engineering from the University of Toronto, Toronto, ON, Canada, in 2002 and the M.S. and Ph.D. degrees in aeronautics and astronautics from the Massachusetts Institute of Technology, Cambridge, in 2004 and 2008, respectively.

He is currently a consultant with McKinsey and Company, Dubai, United Arab Emirates.



Mary L. Cummings (M'03–SM'10) received the B.S. degree in mathematics from the U.S. Naval Academy, Annapolis, MD, in 1988, the M.S. degree in space systems engineering from the Naval Postgraduate School, Monterey, CA, in 1994, and the Ph.D. degree in systems engineering from the University of Virginia, Charlottesville, in 2003.

A Naval Officer and military pilot from 1988 to 1999, she was one of the Navy's first female fighter pilots. She is currently an Associate Professor with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge. Her previous teaching experience includes instructing for the U.S. Navy with the Pennsylvania State University, University Park, and as an Assistant Professor with the Engineering Fundamentals Division, Virginia Polytechnic Institute and State University, Blacksburg. Her research interests include human supervisory control, human-uninhabited vehicle interaction, bounded collaborative human-computer decision making, decision support, information complexity in displays, and the ethical and social impact of technology.