Dynamic Operator Overload: A Model for Predicting Workload During Supervisory Control

Leonard A. Breslow, Daniel Gartenberg, J. Malcolm McCurry, and J. Gregory Trafton

***Abstract*—Crandall *et al.* and Cummings & Mitchell introduced fan-out as a measure of the maximum number of robots a single human operator can supervise in a given single-human-multiple-robot system. Fan-out is based on the time constraints imposed by limitations of the robots and of the supervisor, e.g., limitations in attention. Adapting their work, we introduced a dynamic model of operator overload that predicts failures in supervisory control in real time, based on fluctuations in time constraints and in the supervisor's allocation of attention, as assessed by eye fixations. Operator overload was assessed by damage incurred by unmanned aerial vehicles when they traversed hazard areas. The model generalized well to variants of the baseline task. We then incorporated the model into the system where it predicted in real time, when an operator would fail to prevent vehicle damage and alerted the operator to the threat at those times. These model-based adaptive cues reduced the damage rate by one-half relative to a control condition with no cues.**

***Index Terms*—Cognition, human-robot interaction, multi-robot systems, predictive models, unmanned aerial vehicles.**

I. Introduction

As ROBOTS become cheaper and more autonomous, there is an opportunity to enable one human supervisor to control multiple robots simultaneously. Yet increasing the number of robots that are controlled can hinder operator performance in time-critical supervisory control tasks by increasing operator workload, thereby impacting the operator's attentional resources. Understanding the factors that determine the effectiveness of the overall human–robot system, including factors that affect the cognitive state of the operator, can contribute to the development of adaptive automation that can improve operator performance.

One measure of the number of robots a single operator can supervise at one time is Crandall *et al.*'s fan-out (FO) equation [1]. FO predicts the maximum number of robots that can be monitored by taking into account the amount of time a robot can be neglected before it needs attention ["neglect time" (NT)] in comparison with the amount of time required for an operator to interact with a robot needing attention until it no longer requires attention ["interaction time" (IT)] [1]. The more autonomous the robot, the longer its NT and consequently, the higher the FO, i.e., the number of robots a single operator can control. Similarly, the less IT, higher the FO.

Cummings and Mitchell [2] extended the concept of FO by adding a stronger emphasis on the perceptual and cognitive processes of the operator. Specifically, they included in their FO computation *wait time* (WT) variables, including delays in allocating attention to a vehicle requiring help (WTAA) and delays due to task queuing (WTQ), i.e., allocating time among several vehicles that require attention simultaneously. These WTs constitute time demands in addition to IT that impact FO.

FO is a useful global assessment of a particular task, reflecting the demands the task places upon the operator, thereby facilitating the cognitive engineering design and improving training. We shall explore whether the dynamic variability of performance during the course of a particular task can be predicted by the same, or similar, factors that predict FO for a human–robot system as a whole. Presumably, even where the operator is supervising no more robots than prescribed by the FO equation, there will be moments when events converge to make him/her vulnerable to temporary overloading and, therefore, to error. We shall refer to transitory overload of this sort as *dynamic operator overload.*

We hypothesize that the FO model can be adapted for predicting dynamic operator overload and therefore, provide a basis for preventing operator errors of commission or omission. In Section II, we explain the motivation for our model in terms of predicting operator overload. Our predictive model was created and evaluated over five experiments, described in Section III. The experiments relied on a simulated supervisory control platform in which an operator supervised five homogeneous UAVs. Data from Experiment 1 served to generate the model. Experiment 2 replicated and validated the model by assessing its application to an experimental condition identical to that of Experiment 1. Experiments 3 and 4 assessed the generalizability of the model to different task conditions that were, respectively, relatively easier or more difficult than Experiment 1. Experiment 5 assessed whether the model could predict operator overload in real time by generating cues to warn participants of threats. Specifically, we incorporated the model into the supervisory control simulation to provide real-time cues of upcoming threats when the model predicted damage and compared performance on this system with performance on a system with no cues. Then, in Section IV, we compare dynamic operator overload with fan-out empirically. Finally, in Section V, we conclude our discussion.

Manuscript received April 25, 2013; revised July 30, 2013; accepted October 13, 2013. Date of publication January 2, 2014; date of current version January 16, 2014. This work was supported in part by the Office of Naval Research and Office of the Secretary of Defense to JGT. This paper was recommended by Associate Editor E. Feron of the former IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans (2012 Impact Factor: 2.183).

L. A. Breslow and J. Gregory Trafton are with the Naval Research Laboratory, Washington, DC 20375 USA (e-mail: len.breslow@nrl.navy.mil; greg.trafton@nrl.navy.mil).

D. Gartenberg is with the Department of Psychology, George Mason University, Fairfax, VA 22030 USA (e-mail: dgartenb@masonlive.gmu.edu).

J. Malcolm McCurry is with ITT Exelis, McClean, VA 22102 USA (e-mail: malcolm.mccurry.ctr@nrl.navy.mil).

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/TSMC.2013.2293317

II. From Fan-out to Dynamic Operator Overload

A. Limits of Supervisory Control: Fan-Out

Crandall *et al.* [1] proposed that the maximum number of robots that could be controlled by a single human operator, or FO, could be computed as

$$\mathrm{FO} = \mathrm{NT}/\mathrm{IT} + 1 \tag{1}$$

where NT is the amount of time a robot can be ignored by the operator before its performance drops below some predetermined level, and IT is the amount of time required for the operator to interact with the robot in order to restore the robot's performance to the predetermined acceptable level. This equation defines FO as the maximum number of vehicles an operator can interact with (IT) while another vehicle is running autonomously (NT). "+1" in the equation accounts for the latter, neglected vehicle.

Whereas Crandall *et al.*'s FO model focused on task variables, Cummings and Mitchell [2] extended the model to include the human factors variables, specifically those relating to the overhead of delays, or WTs, WTAA and WTQ, in addition to the duration of direct interaction (IT) with a vehicle requiring attention. These WTs are combined with IT in the denominator of the ratio

$$\mathrm{FO} = (\mathrm{NT}/(\mathrm{IT} + \mathrm{WTAA} + \mathrm{WTQ})) + 1 \tag{2}$$

While FO is a global measure of operator capacity on a task, the amount of operator overload within arbitrary time intervals during the course of a task can be expected to fluctuate as the FO variables drift from their respective mean values. Such fluctuations alter the probability of overload-induced errors over the course of a supervisory control session. Our goal is to instantiate these FO variables, or similar variables, in a model designed to predict when operator load increases enough to cause operators to become overloaded and as a consequence make errors.

In subsequent work, Crandall *et al.* implemented stochastic models of operator–vehicle interactions based on traces of interaction sessions [3], [4]. These models predict the operator's selection of a vehicle to handle and predict vehicle state, based on observed sequences of vehicle states and selections. While this approach was successful in predicting operator performance across task variations, it did not analyze cognitive factors or within-task performance variation, the main foci of the work to be reported here.

While we were concerned with operator load as an impediment to performance, there may be occasions where too *little* operator load can impede performance, as a result of boredom for instance. We do not believe the task we employed was prone to this problem, since it was very demanding, requiring rather continuous attention of the operator.

B. Predicting Operator Overload in a Supervisory Control Task

In this paper, we attempted to predict when an operator supervising multiple UAVs will become overloaded. The simulated control system we used automatically assigned each UAV to a target and determined its initial trajectory toward that target. In addition, there were threats, or hazard areas, that would cause a UAV to be damaged if not avoided. Participants could add waypoints to the trajectory or reassign a UAV to a different target in an effort to prevent damage to a vehicle. Once a UAV arrived at its target, the operator directed it in delivering its payload on the target.

The episodes of interest were path-intersect threat (PIT) events, which start from the moment a vehicle enters on a collision course with a threat and end either at the point in time when the vehicle traverses the threat area, incurring damage, or else at the point in time when the vehicle changes course away from the threat due to the operator's evasive actions. It is clear, when a UAV will traverse a threat area, as a vehicle's trajectory to its target is displayed by a line, which intersects the threat in these cases. However, participants are not specifically alerted to the threat. We assumed that an operator was overloaded when he/she allowed a UAV to incur damage by traversing a threat. We attempted to predict when an operator will fail to prevent a vehicle from taking damage by incorporating variables similar to those in the FO equation within a model designed to predict dynamic operator overload.

As a matter of terminology, we will define the *focal vehicle* of a PIT event to be the vehicle that is on a threat trajectory during that event. Likewise, the threat and target toward which the focal vehicle is heading will be referred to as the focal threat and focal target, respectively, of the PIT event. The focal vehicle, target, and threat will be referred to collectively as the *focal objects*. Vehicles, targets, and hazards other than the focal objects will be referred to as *nonfocal objects, vehicles*, etc. Multiple PIT events may overlap in time, producing one of the main challenges of multiple-vehicle supervision. WTQ represents the amount of time devoted to subtasks related to nonfocal vehicles.

Our interest in predicting damage on a per-event basis motivated a minor change in the FO equation, such that the FO variable NT was replaced by the variable *available time* (AT). AT is the time interval from the start of the PIT event to the expected time of impact with the threat. AT can be determined at the start of a PIT event, based on a vehicle's initial distance from the target, since vehicles in our task moved at a constant speed. During the AT interval, the operator needs to take care of the focal vehicle that is on the threat trajectory, as well as other vehicles requiring attention during that interval, if possible. The number of vehicles that an operator can handle during the AT interval, including the focal vehicle, is the FO. Thus, for the purposes of this paper, we modified FO equation (2) to

$$\mathrm{FO} = \mathrm{AT}/(\mathrm{IT} + \mathrm{WTAA} + \mathrm{WTQ}) \tag{3}$$

where NT is replaced by AT. In addition, the result is not incremented by 1 as it was in (2) because that increment represented a vehicle that can be neglected and there is no vehicle that can necessarily be neglected during a PIT event. Although PIT events often overlapped in time with one another or with payload delivery actions, a detailed analysis revealed that it was rare for a PIT to afford insufficient AT time for vehicle damage to be prevented.

Dynamic operator overload was assessed within each PIT event as the occurrence of damage to the focal vehicle. The variables included in our model predicting damage in a PIT event were operationalized as follows:

1) *Wait Time Attention Allocation (WTAA):* the amount of time it took to recognize that the focal UAV required attention. This duration was operationalized as the duration from the start of a PIT event until the relevant threat was first looked at.

2) *Task Queuing:* represents the allocation of attention to nonfocal objects. Two alternative variables were considered:

a) *Wait Time Queue (WTQ):* As in the FO model, WTQ represents the amount of time spent on manual actions on nonfocal objects.

b) *Wait Queue Fixations (WQF)*: the number of eye fixations on nonfocal objects.

3) *Available Time (AT):* the interval from when a vehicle enters on a collision course with a threat (i.e., the start of a PIT event) until it will make contact with the threat if successful evasive action is not taken. This interval is the amount of time available to the operator to recognize and remedy the threat.

Note that the FO variable IT is not included in our dynamic model of operator overload. In the present context, IT is the time spent on actions resulting in the successful avoidance of damage during a PIT event. Since we are trying to predict the prevention of damage on a per-event basis, activities during IT are clearly partially confounded with what we are trying to predict. Thus, IT does not contribute to our understanding of the processes involved in the occurrence or prevention of damage.

The predictor WQF replaced WTQ, in our model, in part, because it was based on eye fixations, rather than manual actions. Similarly, the predictor WTAA is measured by eye fixations. Eye fixations are a more comprehensive measure of cognitive focus than manual actions, since eye fixations accompany cognitive processes, such as attention allocation, situation assessment, and planning, which can occur with or without concurrent manual actions. In addition, as we shall see, the predictive model based on the eye fixation variable, WQF, was superior to the model based on manual actions, WTQ.

An eye-tracker was used in this paper to record operator's fixations on a computer screen. Eye-trackers are able to measure where an operator is looking (called a fixation) and how long they look at something (called the fixation duration) [5], [6]. Several eye movement measures have been shown to be indicators of cognitive processing [5]–[7]. We used eye fixations as a measure of operator attention allocation. While it is possible to look at a stimulus without attending to it [8], eye movements have been found to correlate with attentional shifts [9]–[11]. As a covert shift of attention seemingly precedes an eye movement to the target of a saccade, eye movements can serve as a direct measure of attention [10]. In addition, the examination of eye movements has been used to predict procedural errors in a manner similar to this paper [12], [13].

Our predictive model of damage in PIT events was computed using logistic regression analysis. Logistic regression computes a multiple linear regression model with a dichotomous outcome variable; a more detailed description can be found in [14]. The dichotomous outcome variable in our analysis of PIT events was the occurrence/avoidance of damage to the focal vehicle. Unlike other classifiers, logistic regression allows one to determine whether or not each of the predictor variables had a statistically significant impact on the overall success of the model, in addition to assessing the model as a whole. Additionally, logistic regression has been used in predictive models of procedural errors in previous research [12], [13].

III. Empirical Evidence

To examine the cognitive processes underlying operator attention and time allocation in a supervisory control task, data were collected from a complex dynamic supervisory control simulation. In the simulation, the participant controlled five semiautonomous, homogenous UAVs. The high-level goal of the simulation was to direct UAVs to specific targets on a map and visually identify key items at the target site in order to deliver the payload on those items. As participants performed the simulation, eye movement and mouse data were recorded.

A critical component to successfully completing the simulation was to prevent UAVs from passing over threat areas, which periodically changed position on the map in an unpredictable manner. If a UAV "hit" (i.e., traversed) a threat area, the UAV took damage and could become incapacitated. Each time a UAV's path intersected a threat area, the operator had to take an explicit action to divert the UAV and prevent damage. Our goal was to develop a model to predict when a vehicle would take damage.

The five experiments that were conducted are outlined in Table I. The data in Experiment 1 were used to generate the predictive model. Experiments 2, 3, and 4 provided validation of the model by testing its generalization to the same procedure as Experiment 1 in Experiment 2, to an easier variant of Experiment 1 (Experiment 3), and to a more difficult variant (Experiment 4). Finally, Experiment 5 tested the application of the model to prevent damage to UAVs by providing users with model-based cues.

A. Method

1) Participants: Participants were George Mason University undergraduate students, who participated for extra credit in a psychology course. All participants had normal or corrected-to-normal vision. Participants were asked to rate how often they played video games on a scale of one (never), two (sometimes), or three (a lot). Participant characteristics are summarized in the last three columns of Table I. As can be seen, participants were sometimes "rejected," i.e., their data were excluded from analysis. Rejection of a participant was most often due to the failure to calibrate a participant on the eye tracker or poor eye tracking validity. Less often, participants' data were excluded due to technical difficulties or experimenter errors.

2) Simulation Description: The supervisory control task, originally designed as the Research Environment for Supervisory Control of Heterogeneous Unmanned Vehicles (RESCHU) [15], [16], was used. RESCHU is a discrete event simulator and it was modified to only include homogenous UAVs moving at a constant speed of 5.2 pixels/s. The interface of the supervisory control simulation, shown in Fig. 1, has three main sections: the map panel, the status panel, and the payload panel. The map

TABLE I
EXPERIMENTAL METHODS, PARTICIPANTS

Experiment	Objective	Method vs. Baseline	# Males: Total, Rejected	# Females: Total, Rejected	Video Game Play: Mean, SD
1. Baseline	Model development	Identical	14, 0	21, 0	1.9, 0.6
2. Replication	Model validation	Identical	15, 2	32, 4	1.8, 0.6
3. Easier	Model validation	Automatic payload delivery	8, 0	25, 0	1.7, 0.6
4. Harder	Model validation	12 s. payload delivery time limit	15, 1	32, 5	1.9, 0.6
5.a.No Cue condition	Model application	Identical	10, 0	12, 0	1.9, 0.6
5.b.Cue condition	Model application	Model-based alerts	9, 0	12, 0	1.9, 0.6

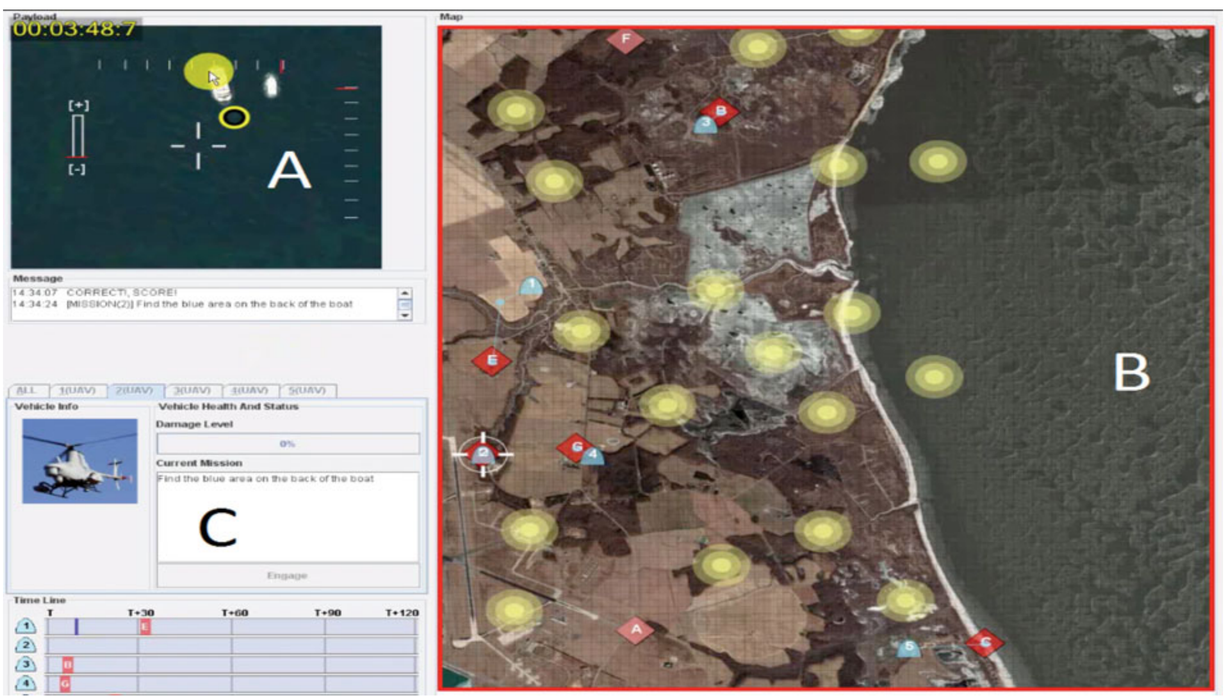

Fig. 1. RESCHU supervisory control simulation, consisting of three panels: A. Payload Panel. B. Map Panel. C. Status Panel.

panel (see Fig. 1, panel B) displays UAVs (blue half ovals), targets (red diamonds) toward which UAVs are moving, and threats (yellow circles) which should be avoided by UAVs. The status panel (see Fig. 1, panel C) shows the status of the UAVs and includes information on vehicle damage, time until the vehicle reaches a waypoint or target, and time remaining in the simulation. The payload panel (see Fig. 1, panel A) is used to acquire a payload target through a visual acquisition task (described below) which is performed after a UAV reaches a target and the target is engaged by the operator.

The operator's high-level goal in the experiments was to monitor UAVs as they proceeded to specific target areas in the map panel, to engage the targets, and to perform a payload visual acquisition task once the UAV had engaged the target. The payload visual acquisition task required the participant to search for the target in the randomly selected photographic image within the payload panel and to mouse-click the target, thereby causing the payload to be delivered. During the visual acquisition payload task, the operator could not interact with the map panel of the interface, even though the vehicles continued to progress toward their targets on the map panel and new PIT events could begin during this interval.

Throughout the session, five UAVs moved along straight line trajectories toward their respective automatically assigned target. There were also 18 threat areas. Every 4 s, one of the 18 threats was randomly selected to change its position to a randomly determined location, with the constraint that the threats could not appear closer than 3° of the visual angle (about 50 pixels) from any UAV. This constraint was imposed to facilitate the eye-tracking system. If the UAV passed through a threat area, it incurred damage. Damage was indicated as a bar in the status panel. The appearance of targets and threats on the simulation map was randomized with the constraint that targets and threats could be no closer than 3° of the visual angle from each other.

If a UAV incurred enough damage, it eventually became incapacitated ("dead"). An incapacitated UAV was colored black and remained immobile on the screen, unavailable for further use. To avoid a threat area, the operator could perform two possible types of action. First, the operator could direct the UAV to a different target. Second, the operator could add, delete, or move waypoints on the UAV's trajectory to route the UAV around the threat, without changing the final destination target.

There were always seven targets present on the map. At the start of the simulation, the UAVs were randomly assigned to different targets; thus, the UAVs might not be directed toward the closest unassigned target. After a target was engaged and the visual acquisition payload task was complete, the UAV was randomly assigned to a new currently unassigned target which again might not be close to it. The simulation was a complex task with multiple events happening in parallel. More than one UAV could be waiting at their respective targets for engagement at the same time and more than one UAV could be on a path intersecting a threat area at a time.

When performing the simulation, participants were scored on their performance, both the number of targets that were engaged correctly and the amount of damage incurred by vehicles. However, no overall score was presented. Participants were instructed to engage as many targets as possible and to prevent damage to vehicles as much as possible. In the version of RESCHU we implemented, the damage to a UAV inflicted by a threat was severe and could permanently incapacitate the UAV. Participants were provided ongoing feedback in the status panel and the map on the amount of damage incurred by vehicles and on the number of incapacitated vehicles.

3) Design and Procedure: Prior to the start of the experiment, participants completed an interactive tutorial that explained all aspects of the simulation. During the tutorial, participants learned the objective of the simulation, how to control the UAVs (changing targets, manipulating waypoints), and how to engage a target and complete the visual acquisition task in the payload panel. Participants were also warned of the dangers of threats and were instructed on how to avoid threats. The tutorial lasted approximately 10 min. After completing the tutorial, the experimenter went over all of the controls with the participant to ensure that the participant understood the task. Participants understood the danger of threat areas and that threat areas could incapacitate a UAV.

After completing the training, participants were seated approximately 66 cm from the computer monitor and were calibrated on the eye tracker. Participants were again instructed to engage as many targets as possible and prevent as much damage as possible. Participants then began the simulation session, which lasted 10 min. When the simulation session ended, participants received feedback on how many vehicles they engaged and total vehicle damage. Then, participants were recalibrated and were run in a second 10-min session with identical procedures to the first session. The data from both sessions were combined in the analyses to be presented.

4) Measures: The data from the supervisory control task were segmented into PIT events. Keystroke and mouse data were collected for each participant. Eye tracking data were collected using an SMI RED eye tracker operating at 250 Hz. A fixation was defined using the dispersion method based on a minimum of 15 eye samples within 60 ms and within 50 pixels (approximately 3° of the visual angle) of each other, calculated in Euclidian distance. Three areas of interest were defined: UAVs, threats, and targets. Other fixations on the map panel and fixations on the payload panel were not analyzed. The eye tracker and the RESCHU simulation were synchronized, such that the simulation sent the eye tracker an update of its state each time its state was updated, i.e., every 500 ms.

We calibrated a participant on the eye tracker until each eye had a visual angle of less than 1°. After ten unsuccessful attempts to calibrate a participant, the participant was not included in the data analysis. Calibration took less than 5 min.

5) Differences in Experimental Methods: The design and procedures of the experiments to be reported were identical to those of the baseline task just described, with the following exceptions:

Experiment 3. Engagement was complete when the participant right clicked on a vehicle and selected the engagement menu item. Unlike the other experiments, there was no need to deliver the payload by performing a visual identification subtask in the payload panel, and, therefore, no interruption of the map panel task.

Experiment 4. A time constraint was imposed on engagement. If the participant failed to initiate payload delivery within 12 s after the UAV reached its target, the UAV was reassigned to a new target without delivering its payload. As in the benchmark condition, payload delivery was accomplished using the payload panel.

Experiment 5. Participants were assigned randomly to either the Cue condition or the control, No Cue condition. Methods in the control condition were identical to the baseline condition. In the Cue condition, a logistic regression model was used in real time to predict whether the participant would fail to prevent a vehicle from hitting a threat area and, if so, to alert the participant of the danger by highlighting the relevant threat. The model reassessed the status of each PIT every 500 ms, when the simulation updated itself. The damage likelihood of each UAV on a path intersect threat course was computed using the dynamic operator overload model, and when the likelihood exceeded the model's threshold value, the relevant threat was highlighted by turning blue (from yellow), and blinking to alert the user of the impending threat.

B. Results

1) Model Development: Experiment 1: Among the 35 participants in the experiment, there were a total of 1999 PIT events, 216 (10.8%) of which ended in damage to UAVs. Mean duration of PIT events was 14 916.2 ms (SD = 16.33). The other main action performed by participants, payload delivery (visual

TABLE II
LOGISTIC REGRESSION TABLE, EXPERIMENT 1

Predictor	β	SE β	Wald X^2	*p*<
Constant	2.17	.31	6.96	.0001
WTAA	.00007	.000009	7.32	.0001
WQF	.11	.007	14.65	.0001
AT	-.00027	.00002	-14.07	.0001

β Values are the coefficients and constant of the model equation. SE β is the standard error of β. Wald X^2 is a metric of the strength of each predictor. p< is the significance level of each predictor.

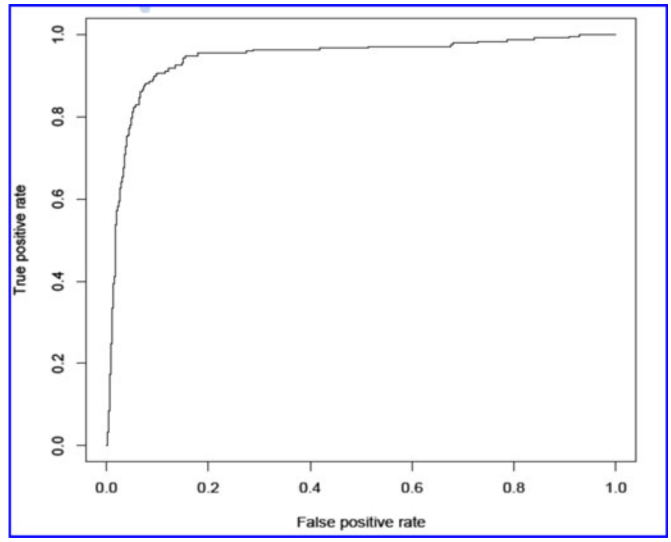

Fig. 2. ROC curve for logistic regression model.

acquisition), had a mean duration of 4800 ms (SD = 400). See Table I for additional information on the participants.

a) Developing a Logistic Regression Model: To create a logistic regression model of the PIT events, the outcomes of damage and no damage were coded as a binary outcome variable for each PIT event. The three predictor variables of interest (WTAA, WQF, AT) were recorded for each PIT event. WTAA and AT were recorded in milliseconds. WQF was an integer representing the quantity of nonfocal fixations during a PIT event. Equation (4) represents the resulting dynamic overload model as a logistic regression equation predicting damage outcomes of PIT events:

$$\text{Predicted Logit of Damage} = 2.17 + (.00007 * \text{WTAA}) + (.11 * \text{WQF}) - (.00027 * \text{AT}) \quad (4)$$

The output of a logistic regression model is a logit; its use in prediction will be explained later. This model was computed based on the final values for each PIT event. In the final experiment, we will examine whether the model is useful for dynamic prediction during PIT events.

The overall logistic regression model was significant, $\chi^2(3) = 240.68$, $p < .0001$. The log odds of damage was significantly related to each of the three predictors ($p < .0001$). The results of the logistic regression model analysis are summarized in Table II. The signs of the β values, representing the coefficients and the constant in the equation, indicate the direction of each predictor's relationship to a damage outcome; thus, all the predictors, other than AT, were positively related to damage. χ^2 Wald is related to the strength of each predictor. WQF and AT were the strongest predictors.

The model fit the data quite well. One measure of fit is the C statistic, which assesses the proportion of all pairs of PIT events with different observed outcomes which the model predicts correctly. The C value of the model was .96, which is excellent as models with C values greater than .80 are considered strong [17]. Thus, for 96% of all relevant pairs of events, the model correctly assigned a higher probability of damage to PIT events that resulted in damage than to events that did not result in damage.

b) Receiver-Operating Characteristic Analysis: Receiver-operating characteristic (ROC) analysis predicts how many damage events from the data were actually predicted by the logistic regression model [18]. Thus, each of the PIT events was classified using the model, and the results were then compared with the actual outcome for that event. In order to classify model outputs according to a binary outcome, such as damage versus no damage, a threshold value must be determined, with model outputs falling above that value classified as damage and those falling below the threshold classified as no damage predictions. An ROC analysis determines the optimal threshold value. Fig. 2 plots the proportion of true positive and false positive classifications for each threshold. The optimal threshold is the one that maximizes true positive classifications and minimizes false positive classifications, and thus corresponds to the upper left-hand point on the curve. The threshold for our model was determined to be 0.26. To classify a PIT event instance, the logit value output by the model equation is converted to a probability using the equation $p = e^{\text{logit}}/(1 + e^{\text{logit}})$ and then compared with the threshold; probabilities greater than the threshold predict an outcome of damage, while probabilities less than the threshold predict no damage.

ROC analysis also provides metrics to evaluate the classifications provided by the model. The area under the ROC curve (AUC) represents the probability that the model will rank a randomly selected positive instance (i.e., damage event) higher than a randomly selected negative instance (i.e., no damage event), and is thus similar to C [18]. Like C, AUC was equal to .96. Finally, ROC analysis provides an overall measure of fit, d', which was equal to 2.65 for the current model, indicative of a highly precise discrimination, since a d' value of 2.0 represents nonrandom discrimination with a 95% probability [19]. The rate of true positive classifications was high (87%), the rate of false positive classifications was low (6%). The results of the ROC analyses, as well as the C score, are displayed in the first row of Table III.

TABLE III
COMPARISON OF LOGISTIC REGRESSION MODELS USING WQF VERSUS WTQ PREDICTORS

Model predictors	C	AUC	d′	True Positive (%)	False Positive (%)
WQF, WTAA, AT	.96	.96	2.65	87	6
WTQ, WTAA, AT	.94	.94	2.43	71	3

We chose to focus on a predictive model that concentrates on eye fixations rather than an alternate predictive model which more closely follows Cummings and Mitchell's FO model, using the variable WTQ in place of WQF. As can be seen in the second row of Table III, the WTQ model displayed a good fit to the data in terms of C and d' scores. However, given the benefits of examining eye movements described earlier, we explored the model described in (4), which relied on the eye movement predictors WQF and WTAA, in the experiments that follow.

c) Discussion: Multiple regression analytic methods, such as logistic regression, assume that the data points are independent, and this assumption is violated in the present model. Each participant contributed on average 57.1 PIT events to the data. As a practical matter, it would be difficult to gather data on 1999 PIT events from that same number of participants. It would be possible in this case to use a mixed-model logistic regression model, but because those models separate fixed and random effects, they can be very difficult to use for prediction because random effects cannot be computed ahead of time for novel participants [20], [21]. The primary concern with not having independent data is that inferences may be incorrect and may not result in accurate generalizations to future datasets. We suspect that the data we collected have exchangeable random variables (future data will behave like past data, regardless of whether it is independent [22]), and the model we construct will generalize to future datasets. The strongest test of this model will occur throughout the rest of this paper where we show that the model does, in fact, generalize to other datasets and can even improve operator performance in real time.

2) Model Validation: Experiments 2–4: The validity of the dynamic overload model was assessed in three experiments: an identical task to Experiment 1 (Experiment 2), an easier task (Experiment 3), and a harder task (Experiment 4). In all experiments, the model used was (4) with the same threshold value as in Experiment 1. Other predictive systems create their models based on an individual; while the model works well on that person, it frequently does not generalize or work well with others. Our approach focuses on building a model that captures perception and cognition at a level that should generalize to anyone and thus does not need a training dataset from each individual.

The damage rate data in Table IV indicate that the manipulations had the intended effects on performance. Relative to the baseline experiment, comparable damage was incurred in the replication, less damage was incurred in the easier task, and more damage was incurred in the harder task. In contrast, payload delivery was relatively unaffected by these manipulations.

TABLE IV
DAMAGE RATE AND PAYLOADS DELIVERED IN EXPERIMENTS

Experiment	Damage rate (%)	Payloads delivered (mean)
1. Baseline	10.8	25.4
2. Replication	10.2	28.4
3. Easier	7.0	NA
4. Harder	13.9	27.0
5.a.No Cue condition	7.5	28.5
5.b.Cue condition	3.7	27.8

TABLE V
MODEL EVALUATION ACROSS EXPERIMENTS

Experiment	C	AUC	d′	True Pos. (%)	False Pos. (%)
1.Baseline	.96	.96	2.65	87	6
2.Replication		.93	2.41	81	7
3. Easier		.95	2.56	86	7
4. Harder		.92	2.16	79	9

While damage rates varied across Experiments 2 to 4, generalization of the dynamic overload model was very good in all three experiments, as indicated in Table V. The results thus support the robustness of the model.

These results also address one potential challenge to our model: It might be possible that participants had different thresholds for what they believed to be an acceptable amount of damage, despite the instructions to try to reduce damage as much as possible and engage as many vehicles as possible. Some participants may have considered a certain amount of damage an acceptable sacrifice in order to engage more vehicles. While individual or systematic biases toward either damage avoidance or engagement cannot be ruled out, the current results suggest that such biases did not affect the generalization of our model. Both when engagement was not performed at all in Experiment 3 and when engagement was given a higher priority in Experiment 4, the dynamic overload model successfully discriminated between events that did versus did not end in damage. Another way to examine the question of different thresholds would be to look at the relationship between the amount of damage and the number of vehicle engagements: if participants were willing to take damage in order to engage more vehicles, there should be a positive correlation between number of engagements and damage. However, across all the experiments, we found a moderate negative relationship between the number of damage instances and the number of payload vehicles engaged ($r = -0.37$, $p < 0.001$), suggesting that the more payload engagements the operator performed the fewer damage instances occurred. This analysis strongly suggests that operators did not, in general, have a strategy to take damage in order to perform more vehicle engagements.

3) Model Application: Experiment 5: The preceding experiments do not directly demonstrate the dynamic overload model's predictive power, since the model was generated and evaluated on data analyzed after the experiments were conducted. In an effort to support the claim that the model is predictive, we applied the model in real time as a means to alert the operator to predicted damage. The cue appeared only in those PIT events where the model predicted that the operator would fail to prevent damage and only once the model first made that prediction. The cue then continued to appear until the end of the PIT event. If using the model to provide cues to the operator results in a decreased rate of vehicle damage, it will provide evidence of the real-time predictiveness of the model.

The model triggered the cue in 61% of the PIT events in the Cue condition. The cue was triggered after an average of 23% of AT (SD = 15%) had elapsed.

As is evident in Table IV, the dynamic cues substantially decreased the rate of damage. Whereas 7.5% of PIT events ended in damage in the No Cue condition, only 3.7% of PIT events ended in damage in the Cue condition. Curiously, the No Cue condition witnessed less damage relative to Experiments 1 and 2, which had comparable procedures. We attribute this anomaly to random variation. In any case, the comparison of the two conditions in Experiment 5 showed that alerting the user to predicted damage via cues reduced the rate of damage by more than half. In addition, in all PIT events ending in damage in the Cue condition, the cue was triggered by the model. That is, there were no cases of damage on PIT events where the cue failed to appear, i.e., no false negative errors. Thus, the model was a strong dynamic predictor of damage and the model-based cue was effective in preventing damage, providing further support for the dynamic operator overload model.

These observations suggest that the damage instances that occurred despite the cue's appearance had a different cause from damage instances predicted by the model. Indeed, the dynamic operator overload model was a poor predictor of damage in the Cue condition. We believe that the remaining instances of damage, not prevented by the cue, were due to *concurrent urgent* PIT events, i.e., nonfocal PIT events that triggered cues on the basis of the damage prediction model (i.e., urgent) that overlapped in time with the focal event (i.e., concurrent). In the Cue condition, there was a mean of 0.84 (SD = 1.09) concurrent urgent nonfocal PIT events. For urgent PIT focal events not ending in damage, the mean rate of concurrent urgent PIT events was similar to this baseline (M = 0.96, SD = 1.13). However, for urgent PIT focal events ending in damage, there were about twice as many concurrent urgent PIT events as the baseline (M = 1.98, SD = 1.33). Thus, damage incurred despite the cue was associated with competition between multiple concurrent urgent PIT events. In many such situations, damage could be avoided only if the focal event was the one first selected by the operator to handle. If not given priority by the operator, such urgent PIT events often ended in damage. Thus, damage that occurred despite the model-generated cue was likely due to task overload that exceeded the operator's capacity.

We should mention that the number of concurrent PIT events in general (i.e., whether urgent or not urgent) was found to not be a good predictor of damage within a logistic regression model in the control, No Cue, condition or in the previous experiments. Consequently, this variable was not included in the dynamic overload model.

TABLE VI
FAN-OUT VALUES FOR ALL EXPERIMENTS

Experiment	AT Fan-out (all)	AT Fan-out (no damage)	NT Fan-out (no damage)
1. Baseline	3.8	4.6	4.8
2. Replication	4.4	5.3	4.9
3. Easier	5.9	6.7	5.1
4. Harder	3.5	4.1	3.9
5.a No Cue condition	4.5	5.3	5.1
5.b Cue condition	4.0	4.2	4.1

Operators controlled five UAVs in all experiments.

We believe that the cue served primarily to encourage attention to the threat, rather than to support cognition of the threat. The RESCHU user interface clearly represents the UAVs' trajectories toward their respective targets graphically by lines. Thus, an upcoming threat could be recognized on a perceptual basis by the visible trajectory's traversal of a threat area. The cue likely drew the users' attention to threats they had not noticed or had forgotten.

In sum, the results of Experiment 5 provided further support for the dynamic operator overload model. The model served as a basis for real-time cues to alert the operator to impending vehicle damage. The cues reduced the damage rate by about half and were never presented unnecessarily in events where there was no damage. Cases where damage occurred despite the cue were characterized by simultaneous cues for more than one threat, suggesting that the overloading was too great for damage from all co-occurring threats to be avoided. Thus, the dynamic operator overload model appears to be a good predictor of supervisor overload.

IV. Fan-Out and Performance Prediction

FO values for each experiment are displayed in Table VI, computed using (3), which we will refer to as "AT fan-out", both based on all PIT events and based only on PIT events where no damage occurred. Note that since FO here is computed only based on PIT events, not on payload events, it does not provide a complete assessment of the demands of the respective tasks. The FO values displayed in Table VI suggest that operators in our experiments were often required to supervise somewhat more vehicles (i.e., five vehicles) than recommended by the FO model. The main exception was the easier task in Experiment 3. In that experiment, the absence of a competing task, payload delivery, reduced the amount of time devoted to competing tasks (i.e., WTQ) during PIT events.

We wished to determine whether AT FO based on AT, as in (3), is similar to FO based on NT, as in Cummings and Mitchell, (2). We computed NT FO as follows:

$$\text{NT fan-out} = (\text{NT/PIT.duration}) + 1 \qquad (5)$$

where NT is time outside of PIT and of payload delivery events, and PIT.duration is the duration of PIT events. PIT.duration served as a substitute for the expression in the denominator of (2), involving WTAA, WTQ, and IT, since the entire PIT event consists of some combination of these durations. During a PIT event, there is always an object that needs attention, namely the focal object. Thus, assuming operator engagement, at every moment in the event, the participant is either working on/attending to the focal object or not working on it. In the former case, the time represents IT, in the latter case it represents WT. WT is either WTQ or WTAA depending on whether the participant is attending to nonfocal objects or not, respectively. Thus, the entire PIT event represents some combination of the denominator variables in (2). Furthermore, no time outside of the PIT event contributes to those variables, with the exception of time spent on the focal vehicle's payload delivery, but that interval is not included in NT.

As Table VI shows, FO values computed based on NT were similar to those based on AT. The principal exception was in Experiment 3, where NT FO was lower than AT FO. The similarity between the two measures of FO supports our interpretation of our logistic regression model as a dynamic version of the Cummings and Mitchell FO model.

A comparison among the first four experiments demonstrates that relative task difficulty was reflected similarly in FO and damage rate. However, the intervention of providing cues in Experiment 5 did not improve FO, even though it radically reduced the damage rate, highlighting an important difference between FO and performance prediction. FO is concerned with having enough time to perform a task, whereas prediction is primarily concerned with what users do with the AT. The cue does not change the amount of time required to perform subtasks but does alert users to direct their efforts to a particular subtask requiring immediate attention. Time intervals where the operator lacked attentional awareness of one problem were not moments of idleness, as the current task is a highly dynamic, time-pressured task in which the operator is continually active. In moments in which the operator has lost awareness of one problem, he/she is generally engaged in another problem. As a result, lost awareness may result in poorer performance without increasing the overall time required for the task.

Another important difference between damage prediction and FO is suggested by a predictive model based on a single variable, i.e., the time remaining to work on the focal threat problem after consuming time on WTAA delay and on working on other, nonfocal objects. This duration represents potential IT

$$\text{potential-IT} = \text{AT} - (\text{WTAA} + \text{WTQ}_1) \quad (6)$$

where WTQ_1 includes both time spent acting on nonfocal objects and time spent fixating such objects and where both of those durations are calculated so as not to overlap with the WTAA interval, i.e., the initial duration of the PIT before the focal threat is first fixated. As Table VII shows, the single-variable model displays comparable quality to the dynamic operator overload model, defined by (4). Furthermore, the single-variable model adheres more closely to the principle underlying the FO equation, as it is based solely on an estimate of the time remaining to work on the focal problem after deducting all WTs from the AT.

TABLE VII
COMPARISON OF THE DYNAMIC OPERATOR OVERLOAD MODEL AND A MODEL BASED ON POTENTIAL-IT, EXPERIMENT 1 DATA

Model	C	AUC	d′	True Positive	False Positive
Dynamic Operator overload Model	.96	.96	2.65	87%	6%
Potential-IT Model	.97	.97	2.83	82%	3%

However, this model can be shown to be much less useful than the original model in terms of the timeliness of the prediction. In the Cue experiment (Experiment 5), the dynamic operator overload model produced a warning signal after only 23% of the AT had elapsed, on average. In contrast, solving the single-variable model's logistic regression equation for the value required to exceed the model's threshold shows that the model would not produce a warning signal until 90% of the AT had elapsed. That is too late to be useful to the operator. FOs based on (3) under procedures similar to the Experiment 5 control condition (i.e., Experiments 1 and 2) are generally between 4 and 5 (see Table VI). Thus, approximately 20–25% of AT is required to take care of each vehicle. Warnings provided by the potential-IT model, when only 10% of AT remains, are clearly insufficient, whereas the warnings provided by the dynamic operator overload model are more than adequate, allowing 77% of the AT to take care of the vehicle needing attention.

These observations highlight one key difference between FO and predictive models. FO models analyze performance on a task globally and, therefore, are not aimed at within-task prediction, beyond a global prediction of how many vehicles an operator will generally be able to supervise in a given task. The dynamic operator overload model, in contrast, is useful for prediction of performance *during* a task session.

However, the comparison of the two logistic regression models demonstrates that even logistic regression does not always provide useful predictions of real-time performance. Logistic regression makes no distinction among the relative temporal position of data points within an interval (e.g., a PIT event) and, therefore, requires theory-based selection of possible predictors by the researcher in order to contribute to timely prediction. The variable potential-IT diminishes progressively from the start to the end of the PIT event. In contrast, in the dynamic operator overload model, the value of one predictor, AT, is known at the start of the PIT event. A second predictor, WTAA is an interval that begins at the start of the event and that usually ends well before the event is finished. Only the third predictor, WQF (or WTQ), grows progressively throughout the course of the event. As a result, the dynamic operator overload model provides a better basis for an alert system than the potential-IT (6) model.

In addition, a predictor to be useful must be theoretically meaningful, as illustrated by the potential-IT model, which is based only on time remaining to perform a task. It is trivial that a participant who never performs an action will run out of sufficient time to do so shortly before the deadline. To paraphrase a well-known saying, it is always darkest before nightfall. It is more useful from a theoretical and practical perspective to know that the operator's failure to notice a problem and the operator's preoccupation with other objects and activities predict the failure to correct the problem. More generally, high scores on typical criteria to assess logistic regression models (C, d', etc.) are not sufficient to guarantee that a model is theoretically or practically useful.

V. Conclusion

The FO models of Cummings and Mitchell, and Crandall *et al.* were designed to estimate the number of UAVs a single operator can supervise. Their estimates are based on time intervals, including the length of time a vehicle may be ignored before its performance degrades below a specified threshold (NT), the time required to bring a vehicle's performance back up above the threshold (IT), and delays between those two intervals due to loss of attentional awareness (WTAA) and due to time spent on higher-priority tasks (WTQ).

We explored the relationship between *system-focused* FO, on which Cummings and Crandall focused, and *dynamic operator overload*, which varies over the course of operator–system interaction. Even when an operator is required to supervise no more vehicles than dictated by the system-focused FO model, there may be moments when dynamic task demands converge to overload the operator, resulting in errors. It is the goal of a dynamic operator overload model to predict such situations of transitory overload.

In this paper, the dynamic operator overload model was developed to predict the quality of ongoing performance of novice operators engaged in a simulation task in which they supervised five UAVs, attempting to keep them from incurring damage by traversing threat areas while directing the vehicles to deliver payloads on assigned targets. We developed logistic regression models to predict vehicle damage based on ongoing operator behaviors and attention as assessed by operator eye movements. Our models took the variables that figure in system-focused FO models as their starting point.

The FO variables WTAA and WTQ, together with the variable AT (substituted for NT for task-specific reasons), yielded a model that was highly predictive of damage occurrences in PIT events.

More predictive still was a model that substituted number of fixations on nonfocal objects (WFQ) in place of time spent acting on nonfocal objects (WTQ). The superiority of the model with the WFQ variable over the model using WTQ may be due to the fact that fixations are a more comprehensive measure than manual actions and are more sensitive to individual differences in visual and/or core processing speed. Fixations on objects occur during manual actions on those objects (i.e., as in WTQ) as well as during cognitive activity in the absence of overt action, such as scanning and decision making. The WFQ-based model was chosen for further examination in this paper.

The model's parameters were generated from the data in Experiment 1. The model was then replicated in Experiment 2, generalized to an easier task and a harder task in Experiments 3 and 4, respectively, demonstrating the model's robustness. The model was then applied in Experiment 5, where the model initiated cues that alerted the user to impending damage.

This paper also pointed to the importance of attention in predicting performance. Two of the three variables in the dynamic operator overload model, WTAA and WFQ, were based on eye fixation data. WTAA, the time it took for the operator to first fixate on the relevant threat of a path-intersects threat event, is clearly related to attention. WFQ, the number of fixations on nonfocal objects may reflect attention in part. The success of the alert cue in Experiment 5 in greatly reducing damage rate likewise suggests the importance of attention to task performance.

The success of the model-based cues in Experiment 5 also provided evidence of the ability of the dynamic operator overload model to predict damage in real time. The cues reduced the rate of damage by about half. What is more, no damage occurred in the absence of a cue. The success of the cues also points to the potential practical application of the dynamic operator overload model.

This paper also sheds light on considerations involved in developing a model that provides *timely* feedback, as far in advance as possible. We demonstrated that a model that is a highly discriminating classifier may not be able to make a decision until late in an event, at which point a warning likely comes too late to be useful. In contrast, our model, when used as the basis for user cues, was able to alert the user of a threat after only 23% of the event had passed. We argued that the timeliness of the cue was due to the model's reliance on factors most of which were available at, or soon after, the start of the event.

The main conclusion of this paper is that a system-focused FO model may be adapted to produce a dynamic operator overload model. Both models rely on the operator's allocation of AT to competing subtasks. Whereas system-focused FO is a global assessment of a task, the dynamic operator overload model allows for variations in both AT and the number of competing time demands during the course of a task, the focus of dynamic operator overload. Some of this variation is imposed by the environment, outside of the operator's control, whereas other variations in time allocation are due to the operator's attentional awareness and the operator's decision and planning skills. Both types of variation contributed to the dynamic operator overload model presented here.

Acknowledgment

The authors would like thank M. Cummings and the Human Automation Laboratory at Massachusetts Institute of Technology for providing us with RESCHU. The authors also thank J. Park and G. Picci for providing assistance in conducting the experiments.

REFERENCES

[1] J. W. Crandall, M. A. Goodrich, D. R. Olsen, and C. W. Nielsen, "Validating human-robot interaction schemes in multitasking environments," *IEEE Syst., Man, Cybern. A, Syst. Humans*, vol. 35, no. 4, pp. 438–449, Jul. 2005.

[2] M. L. Cummings and P. J. Mitchell, "Predicting controller capacity in supervisory control of multiple UAVs," *IEEE Syst., Man, Cybern. A, Syst. Humans*, vol. 38, no. 2, pp. 451–460, Mar. 2008.

[3] J. W. Crandall, M. L. Cummings, and C. E. Nehme, "A predictive model for human-unmanned vehicle systems," *J. Aerosp. Comput. Inf. Commun.*, vol. 6, pp. 585–603, 2009.

[4] J. W. Crandall, M. L. Cummings, M. Della Penna, and P. M. A. de Jong, "Computing the effects of operator attention allocation in human control of multiple robots," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 41, no. 3, pp. 385–397, May 2011.

[5] K. Rayner and R. K. Morris, "Do eye-movements reflect higher order processes in reading?" in *From Eye to Mind. Information Acquisition in Perception, Search, and Reading*, R. Groner, D. d'Ydewalle, and R. Parham, Eds. Amsterdam, The Netherlands: North-Holland, 1990, pp. 191–204.

[6] K. Rayner, "Eye movements in reading and information processing: 20 years of research," *Psychological Bulletin*, vol. 85, pp. 618–660, 1998.

[7] M. A. Just and P. A. Carpenter, "Eye fixations and cognitive processes," *Cognitive Psychology*, vol. 8, pp. 441–480, 1976.

[8] I. D. Brown, "A review of the 'looked-but-failed-to-see' accident causation factor; road safety research," Dept. for Transport, U.K., London, U.K., Rep. no. 60, 2002.

[9] J. M. Henderson and A. Hollingworth, "The role of fixation position in detecting scene changes across saccades," *Psychological Sci.*, vol. 10, pp. 438–443, 1999.

[10] M. S. Peterson, A. F. Krame, and D. E. Irwin, "Covert shifts of attention precede involuntary eye movements," *Perception Psychophysics*, vol. 66, pp. 398–405, 2004.

[11] K. Rayner, G. W. McConkie, and S. Ehrlich, "Eye movements and integrating information across fixations," *J. Exp. Psychology: Human Perception Perform.*, vol. 4, pp. 529–544, 1978.

[12] R. M. Ratwani and J. G. Trafton, "A generalized model for predicting postcompletion errors," *Topics Cognitive Sci.*, vol. 2, pp. 154–167, 2010.

[13] R. M. Ratwani, J. M. McCurry, and J. G. Trafton, "Single operator, multiple robots: An eye movement based theoretic model of operator situation awareness," in *Proc. 5th ACM/IEEE Int. Conf. Human–Robot Interaction*, 2010, pp. 235–242.

[14] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *J. Educational Res.*, vol. 96, pp. 3–14, 2002.

[15] Y. Boussemart and M. L. Cummings, "Behavioral recognition and prediction of an operator supervising multiple heterogeneous unmanned vehicles," *Humans Operating Unmanned Syst.*, Brest, France, Sep. 2008.

[16] C. Nehme, "Modeling human supervisory control in heterogeneous unmanned vehicle systems," Ph.D. dissertation, Dept. Aeronautics Astronautics, Massachusetts Inst. Technol., Cambridge, MA, USA, 2009.

[17] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. New York, NY, USA: Wiley, 2000.

[18] T. Fawcett, "An introduction to ROC analysis," *Pattern Recog. Lett.*, vol. 27, pp. 861–874, 2006.

[19] J. A. Swets, *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1996.

[20] D. Bates, (2010) lme4: Mixed-effects modeling with R. [Online]. Available: http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf

[21] D. Bates and M. Maechler. (2009) lme4: Linear mixed-effects models using S4 classes. [Online]. Available: http://CRAN.R-project.org/package = lme4

[22] G. J. Szekely and J. G. Kerns, "De Finetti's theorem for abstract finite exchangeable sequences," *J. Theoretical Probability*, vol. 19, pp. 589–608, 2006.

Leonard A. Breslow received the B.A. degrees in philosophy and psychology from Yale University, New Haven, CT, USA, in 1973 and the Ph.D. degree in psychology from the University of California, Berkeley, Berkeley, CA, USA, in 1981.

He is currently a Cognitive Scientist with the Naval Research Laboratory, Washington, DC, USA. His research focuses on color visualizations, predicting and preventing errors, and supervisory control.

Daniel Gartenberg received the B.S. degree in psychology and English from the University of Wisconsin-Madison, Madison, WI, USA, in 2009, and the M.A. degree in psychology from George Mason University, Fairfax, VA, USA, in 2011, where he is currently a graduate student working toward the Ph.D. degree in psychology.

J. Malcolm McCurry received the B.A. degree in psychology from George Mason University, Fairfax, VA, USA, in 2005.

He is currently with ITT Exelis, McLean, VA, USA, focusing on vigilance, errors, interruptions, and supervisory control.

J. Gregory Trafton received the B.S. degrees in computer science and psychology from Trinity University, San Antonio, TX, USA, in 1989, and the M.S. and Ph.D. degrees in psychology from Princeton University, Princeton, NJ, USA, in 1994.

He is currently a Cognitive Scientist with the Naval Research Laboratory, Washington, DC, USA. His research focuses on human–robot interaction, predicting and preventing errors, and supervisory control.