Assessing Operator Strategies for Adjusting Replan Alerts in Controlling Multiple Unmanned Vehicles

Citation: Maere, P. C., Clare, A. S., Cummings, M. L., "Assessing Operator Strategies for Adjusting Replan Alerts in Controlling Multiple Unmanned Vehicles," 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human- Machine Systems, Valenciennes, France, Aug. 2010.

As Published: http://www.univ-valenciennes.fr/IFACHMS2010/abstact/13-A-Assessing_operator_strategies_for_adjusting_replan_alerts_in_controlling_multiple_unmanned_vehicles_Pierre_Cp_Maere.pdf

Publisher: International Federation of Automatic Control

Persistent URL: http://hdl.handle.net/1721.1/66241

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Maere, P. C., Clare, A. S., Cummings, M. L., "Assessing Operator Strategies for Adjusting Replan Alerts in Controlling Multiple Unmanned Vehicles," 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France, Aug. 2010. http://www.univ-valenciennes.fr/IFACHMS2010/

Assessing Operator Strategies for Adjusting Replan Alerts in Controlling Multiple Unmanned Vehicles

Pierre C. P. Maere*. Andrew S. Clare.**
Mary L. Cummings.***

** Delft University of Technology, Delft, Netherlands*
(Tel: 0031-628-429-530; e-mail: p.c.p.maere@student.tudelft.nl).
*** Aeronautics and Astronautics Department, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, (e-mail: aclare@mit.edu)*
**** Aeronautics and Astronautics Department, Massachusetts Institute of Technology, Cambridge, MA 02139 USA, (e-mail: missyc@mit.edu)}*

Abstract: This study examined the impact of allowing an operator to adjust the rate of prompts to view automation-generated plans on operator performance and workload when supervising a decentralized network of heterogeneous unmanned vehicles. *Background*: Future unmanned vehicles systems will invert the operator-to-vehicle ratio so that one operator can control multiple vehicles with different capabilities, connected through a decentralized network. A previous experiment showed that higher rates of replan prompting led to higher workload and lower system performance. Poor performance was associated with a lack of operator consensus for when to accept the automation's suggested prompts for new plan consideration. *Method*: Three initial rates of replanning were tested on an existing, multiple unmanned vehicle simulation environment that leverages decentralized algorithms for vehicle routing and task allocation, in conjunction with human supervision. Operators were provided with the ability to adjust the rate of replanning. *Results*: The majority of the operators chose to adjust the rate at which they were prompted to replan. Operators favored particular replan intervals, no matter which initial replan interval they started at. It was found that different initial replan intervals produced differences in mission performance. In addition, increasing amounts of replanning caused the system to destroy more targets but do a poorer job at tracking targets. *Conclusion*: Operators have preferences for the rate at which they prefer to view automation-generated plans. Allowing operators to institute these preferences influenced the overall mission performance. Further research is necessary to determine the full impact of the operators' strategies for changing the replan intervals on net mission performance. *Application*: Future unmanned vehicles systems designs should incorporate the flexibility to allow operators to adjust the frequency at which the automation generates new plans for approval.

Keywords: Automated guided vehicles, Automation, Decentralised systems, Human-machine interface, Human supervisory control, Mental workload, Multimachine, Routing algorithms.

1. INTRODUCTION

A future concept of operations for controlling unmanned vehicles is one of a single, forward-deployed soldier supervising multiple, heterogeneous (air, sea, land) unmanned vehicles (Naval Studies Board, 2005). In order to achieve this concept of operations, significant collaborative autonomy will have to be embedded within and across these teams of vehicles, so that the vehicles can execute basic operational and navigational tasks autonomously (Cummings et al., 2007). Operators will supervise these vehicles by providing high level direction to achieve mission goals. They will need to comprehend a large amount of information while under time pressure to make effective decisions in a dynamic environment. They will be assisted by automated planners to reduce workload. As a result, human management of the automated planner is crucial, as auto-planners do not always generate accurate solutions. Though fast and able to handle complex computation better than humans, computer optimization algorithms are notoriously "brittle" in that they can only take into account those quantifiable variables identified in the design stages that were deemed to be critical (Smith et al., 1997).

In a previous experiment, the impact of increasing automation replanning rates on operator performance and workload was examined (Clare et al., 2010). The operator was prompted to replan at various intervals, but could choose to replan whenever he or she desired. When replanning, the operator was presented with plans created by the automated planner, which he or she could accept, reject, or attempt to modify manually. Results showed that the rate of replanning by the human operator had a significant impact on workload and performance, with higher replanning rates resulting in degraded performance. Results from this experiment also showed that operators who collaborated with the automated planner, labeled "Consenters", had significantly higher performance and lower workload than those who ignored automation requests for replan consideration (Cummings et al., 2010).

The experiment described in this paper builds on the previous experiment by allowing operators to set the rate of replan prompting. The purpose of this research is to see if there is a replanning rate that human operators prefer, and whether there is an effect on performance. In the following sections, we discuss the testbed created that allows for this investigation, as well as the methodology to investigate operator strategies for collaborative decision making in a dynamic environment.

2. EXPERIMENTAL TEST BED

The mission of interest is search, track, and destruction of enemy targets. The objective of the operator is to command multiple, heterogeneous unmanned vehicles for the purpose of searching the area of responsibility for new targets, tracking targets, and approving weapons launch. Once a target is found, it is designated as hostile, unknown, or friendly, and given a priority level by the user. Hostile targets are tracked by one or more of the vehicles until they are destroyed by a weaponized UAV. Operators must approve all weapon launches. Unknown targets are revisited as often as possible, tracking target movement.

Provided with intelligence via a chat box, the operator has the ability to re-designate unknown targets or create search tasks for emergent targets. The primary interface used by the operator is a map display, shown in Fig 1. The operator is assisted by an automated planner in scheduling the search tasks and target tracking assignments to be completed by the UxVs.

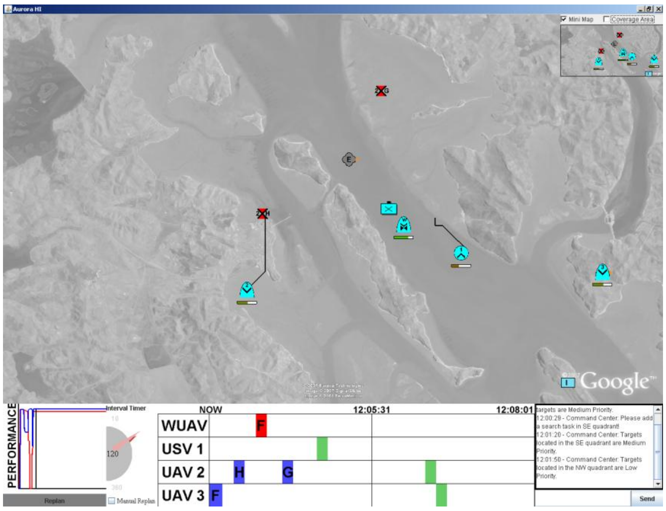

Fig 1: Map Display

In order to aid the operator in understanding the current state of the UxVs and their progress towards mission goals, a decision support interface, called the Schedule Comparison Tool, was developed, shown in Fig 2. Details of the interface design and usability testing can be found in Fisher (2008).

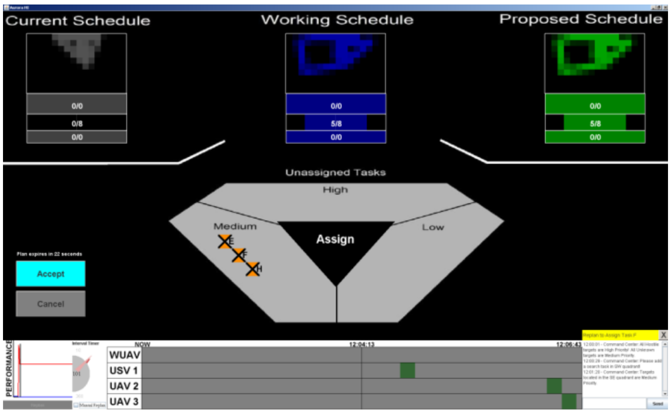

Fig 2: Schedule Comparison Tool

Given previous results that showed that at least one-third of operators ignored preset automation replanning intervals, a new component was added to the map display, as shown in Fig 3. The operator can adjust the replan prompting interval by using the Replan Interval Dial. The dial can be set to a preferred value, as long as it is between the two boundary values, 1 and 360 seconds. In addition, automated replanning intervals could be disabled by selecting the manual replan option.

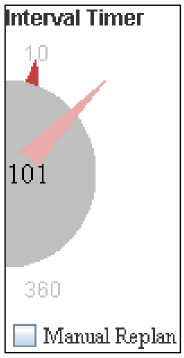

Fig 3: Replan Interval Dial

3. METHODOLOGY

In futuristic multiple unmanned vehicle settings, the rate at which a human operator must confirm or alter plans will likely have a significant impact on operator workload. A previous experiment (Clare et al., 2010) showed that higher rates of replanning caused higher workload and degraded system performance. Operators were prompted to replan at three different intervals: 30 seconds, 45 seconds, and 120 seconds. The prompt was given through the green illumination of the replan button and an aural replan alert when a schedule was available that was different from the current schedule. Although the automation could generate plans as rapidly on the order of seconds, operators would only be notified of a new plan at the intervals of 30 seconds, 45 seconds, and 120 seconds, depending on which scenario was being tested.

Additionally, operators who consented to the rate of replanning based on the prompting interval had significantly lower workload and higher system performance (Cummings et al., 2010). The label Consenter did not mean that an operator necessarily agreed with the plan proposed by the automated planner, but only that the operator agreed to view the new plan in the Schedule Comparison Tool. The group of operators that was labeled “Dissenters” had nearly identical self-imposed replanning rates in that they queried the automation for a new plan approximately every 30 seconds (Cummings et al., 2010).

The experiment detailed here provides operators with the ability to set the rate of replan prompting. The aim is to determine whether there is a replanning rate that human operators prefer and whether there is an effect on operator workload, system performance, or subjective operator assessment of the system. In this experiment, the independent variable was the initial replan prompting interval. The three levels for the independent variable were 30, 45, and 120 seconds, identical to the replan prompting rates from the previous experiment (Clare et al., 2010). The key difference is that operators could change this prompting interval to any interval between 1 and 360 seconds. Operators also had the option to turn of the replan prompt, so that there would be no notification of when the automation had a new plan for the operator to evaluate.

The dependent variables included objective workload metrics, mission performance metrics, and subjective operator ratings of performance, confidence, and workload. We elected to measure workload via a utilization metric (i.e., percent busy time) because utilization has proven to be sensitive to changes in workload in similar multiple tasking, time-pressured scenarios (Cummings and Guerlain, 2007, Cummings and Nehme, 2009). Operators were considered "busy" when performing one of the following tasks: creating search tasks to specify locations on the map where UxVs must search for targets; identifying targets by looking at the imagery and designating a target type and priority level; approving weapons launches on hostile targets; chat messaging with the command center on intelligence information; and replanning in the SCT. The mission performance dependent variables included the percentage of area covered, the percentage of hidden targets found, the percentage of time that targets were tracked, and the percentage of hostile targets destroyed. Finally, a survey was presented to the operator after each of the three scenarios to gather subjective self-ratings of performance, confidence, and workload on a 1-5 Likert scale.

This experiment was conducted using two Dell 17" flat panel monitors operated at 1280 x 1024 pixels and a 32-bit color resolution. The primary monitor displayed the testbed and the secondary monitor showed a legend of the symbols used in the system. The workstation was a Dell Dimension DM051 with an Intel Pentium D 2.80 GHz processor and a NVIDIA GeForce 7300 LE graphics card. System audio was provided using standard headphones that were worn by each participant during the experiment. All data regarding the human participant's interactions with the system for controlling the simulated UxVs was recorded automatically by the system.

In order to familiarize each subject with the interfaces in Figures 1 and 2, a self-paced, slide-based tutorial was provided, which typically took subjects approximately twenty minutes to complete. Then, subjects had a ten-minute practice session during which the experimenter walked the subject through all the necessary functions to use the interface and to develop working plans before accepting them. Each subject was given the opportunity to ask the experimenter questions regarding the interface and mission during the tutorial and practice session.

The actual experiment for each subject consisted of three 15 minute sessions. The three possible initial intervals between automation-generated replan proposals were 30 seconds, 45 seconds and 120 seconds. Each subject experienced each of these initial rates in a counterbalanced and randomized order to prevent learning effects. Subjects were able to change the intervals between proposals using the Replan Interval Dial. Each scenario was different, but similar in difficulty. The interface recorded all operator actions.

Subjects were selected from a sample population similar to that which the military is interested in for the types of operations simulated by this interface. The subject population consisted of twenty-nine subjects: 20 men and 9 women. Ages ranged from 18 to 31 years with a mean of 23.58 years and standard deviation of 3.57 years. All subjects had previous experience with this simulation testbed without the replan dial. 16 subjects had participated in a previous experiment with this testbed at least a month prior, and were labeled "Experienced", while 13 received equivalent practice time with the interface (without the replan dial) immediately prior to this experiment, and were labeled "Inexperienced." About a third of subjects had military experience (ROTC, Air Force Academy, or Active Duty). Each subject filled out a demographic survey prior to the experiment that included age, gender, occupation, military experience, comfort level with computers, and video gaming experience.

4. RESULTS AND DISCUSSION

4.1 Replan Dial Strategy

An analysis of the operators' strategies was conducted based on experimental data. First, we analyzed how much operators used the replan dial to modify the replan prompting rate. Of the 87 test trials, the replan dial was utilized in 54 of them, or 62% of the trials. Two of the 29 operators never changed the prompting interval in any of the 3 scenarios. Another 9 operators only made changes to the prompting interval in 1 of the 3 scenarios. Of those 9 operators, 7 made a change to the prompting interval when the initial prompting interval was 120 seconds. This reflects a finding from the previous experiment (Cummings et al., 2010), where the highest performing group generally followed the automation's suggestions for plan consideration for the 30s and 45s replan intervals, but generally could not wait the full 120s. In this experiment, 7 of the 29 operators left the replan dial untouched, except in the 120s case, where they lowered the prompting interval.

For the 30 second initial replan prompting rate, 12 of the 29 operators never changed the prompting interval. For the 45 second initial replan prompting interval, 13 operators never changed the prompting interval. For the 120 second initial replan prompting interval, however, only 8 operators never changed the prompting interval. This aligns with the previous insight that operators were willing to consent to replan at the 30s and 45s intervals, but not at the 120s interval. This information is summarized in Table 1.

Table 1. Operators who did not use Replan Dial

	30s Initial Replan Interval	45s Initial Replan Interval	120s Initial Replan Interval
Number of Operators who Never Adjusted the Replan Dial	12	13	8

For operators who did use the replan dial, the average number of replan dial changes was evaluated. For operators who made at least one change, the average number of changes to the replan dial was 2.44. The breakdown for the average number of changes for each of the initial replan intervals is shown in Table 2. The average and final replan prompting intervals (for only those operators that made a change from the initial interval) are also shown in Table 2. A wide range of final values was chosen by the operators, with values typically higher than the initial value for the 30 and 45 second intervals, and lower than the initial value for the 120 second interval scenario. Finally, one subject chose to switch the replan prompt to "off" as the last replan dial setting for all 3 of the trials, while one other subject switched the prompt to "off" at the 30 second initial replan interval. This data is summarized in Table 2.

Table 2. Replan Prompting Strategy of Operators Who did Use the Dial

	30s Initial Replan Interval	45s Initial Replan Interval	120s Initial Replan Interval
Number of Changes to Replan Prompting Interval	2.41 ± 1.8	2.75 ± 2.8	2.24 ± 2.1
Average Replan Prompting Interval	62.46 ± 40.1s	71.19 ± 34.7s	91.69 ± 39.1s
Final Replan Prompting Interval	92.35 ± 88.1s	104.75 ± 90.4s	79.14 ± 78.2s
Number of Operators who Ended the Scenario with Replan Prompting Off	2	1	1

Upon further analysis, it appeared that operators tended to favor the same replan prompting interval value, no matter which initial value they began at. Operators were partitioned into 3 Replan Prompt Categories: 1) Low Replan Values, 2) Medium Replan Values, and 3) High Replan Values. The partition was based on the preference of the operators for any of these replan values. For example, the operators in category 1 favored low replan values and thus faster replan prompting. These three populations were not evenly split (65%, 14%, and 21%, respectively), with a greater number of subjects favoring shorter replan prompting intervals. The average number of changes to the replan dial for each group was 2.7, 2.1, 2.1, respectively. The Low Replan Value group made the most changes to the Replan Dial. Table 3 shows the intended replan rates and the final replan prompting rate adopted by participants in the three categories. It should be noted that all 4 subjects in the Medium Replan Values group, when starting with an initial 120s replan interval, ended the scenario with the Replan dial set to 120 seconds.

Table 3. Replan Prompt Category and Final Replan Prompting Intervals

	30s Initial Replan Interval	45s Initial Replan Interval	120s Initial Replan Interval
Low Replan Values	35.6 ± 15.7s	41.7 ± 6.6s	59.3 ± 32.3s
Medium Replan Values	53.3 ± 51.4s	105.3 ± 51.2s	120.0 ± 0s
High Replan Values	173.3 ± 101.7s	174.7 ± 105.3s	169.3 ± 103.7s

4.2 Workload Metrics

Descriptive statistics were calculated for the workload metric, utilization, which measured the percent busy time of operators during their missions. A trend was found for experienced operators to have higher utilization than inexperienced operators, as shown in Table 4. A Mann-Whitney Independent test confirmed that there was a marginally significant difference in utilization between experienced and inexperienced operators, Z = -1.870, p = 0.061. Operators who had used the simulation testbed in a previous experiment likely tried to guide the automation more than inexperienced operators who may have intervened less. This trend held true for both operators who modified the replan prompting interval and for operators who did not use the Replan Dial.

Table 4. Utilization Across Experience Levels

Experience Level	Initial Replan Interval	Mean (%)	Std Dev (%)
Inexperienced	30	35.8%	8.15%
	45	36.2%	5.51%
	120	33.3%	6.48%
Experienced	30	39.9%	7.83%
	45	37.4%	6.95%
	120	36.3%	8.49%

Comparing the workload of the 3 Replan Prompt Categories, we found no significant differences between the three Replan Values Group. Table 5 summarizes the key statistics.

Table 5. Workload Metrics Across Replan Prompt Categories

Replan Prompt Category	Mean (%)	Std Dev (%)
Low Replan Values	36.2%	6.53%
Medium Replan Values	40.7%	6.63%
High Replan Values	35.9%	9.91%

4.3 Performance Metrics

The four overall mission performance metrics were percentage of area coverage, percentage of targets found, percentage of time that targets were tracked, and number of hostile targets destroyed. In terms of the effect of experience on system performance, a trend was found towards more experienced users having higher area coverage. A Mann-

Whitney Independent test confirmed that there was a marginally significant difference in area coverage between experienced and inexperienced operators, Z = -1.939, p = 0.052. These results are shown in Table 6.

Significant differences were found in mission performance based on initial replan interval. The Kruskal-Wallis omnibus test showed a significant difference in targets found based on initial replan interval, $\chi 2(2, N=87) = 18.368$, $p< 0.001$, with the lowest number of targets found at the 30 second initial interval. A similar trend was found for the percentage of time that targets were tracked, $\chi 2(2, N=87) = 17.947$, $p< 0.001$, with the lowest tracking times at the 30 second initial interval. Finally, there was a significant difference in the number of hostiles destroyed based on initial interval, $\chi 2(2, N=87) = 26.936$, $p< 0.001$, with the most number of hostiles destroyed at the 45 second interval. These findings support previous experimental results showing that rapid rates of replanning, at the 30 second interval, can lead to lower system performance (Clare et al., 2010).

We compared the performance of operators based on Average Replan Interval. Results showed that there was a marginally significant correlation between increasing average replan interval and increasing time that targets were tracked, $\rho = 0.197$, $p = 0.067$. However, it was shown that there was a significant correlation between increasing average replan interval and decreasing number of hostiles destroyed, $\rho = -0.398$, $p = 0.000$. Increasing amounts of replanning caused the system to destroy more targets but to do a poorer job of tracking targets. This indicates that the addition of the capability for operators to change the rate at which the automation presents new plans had a significant effect on performance. Finally, there were no significant differences in performance based on Final Replan Interval or Replan Prompt Categories.

Table 6. Performance Metrics Across Experience Levels

Metric	Experience Level	Initial Replan Interval	Mean	Std Dev
% Area Coverage	Inexperienced	30	44.8%	9.81%
		45	46.7%	8.80%
		120	45.4%	8.43%
	Experienced	30	47.7%	9.50%
		45	52.9%	9.73%
		120	48.2%	8.96%
% Targets Found	Inexperienced	30	70.8%	11.65%
		45	79.2%	9.96%
		120	82.5%	9.65%
	Experienced	30	65.9%	10.04%
		45	79.4%	14.78%
		120	80.6%	12.98%
% Time Targets Tracked	Inexperienced	30	87.0%	9.50%
		45	96.6%	3.50%
		120	95.9%	3.37%
	Experienced	30	90.5%	7.28%
		45	96.1%	3.58%
		120	96.7%	3.04%
Hostiles Destroyed	Inexperienced	30	3.2	1.19
		45	3.8	1.03
		120	2.1	0.90
	Experienced	30	3.1	0.90
		45	3.5	1.00
		120	2.1	0.90

4.4 Subjective Self-Rating Metrics

A survey was provided at the end of each mission asking the participant for a subjective rating of his or her workload and confidence on a Likert scale from 1-5. There was a marginally significant correlation between increasing average replan interval with decreasing workload self-ratings, $\rho = -0.192$, $p = 0.075$. This is to be expected, as the longer the interval between replanning, the less work the operator should be performing.

There were no significant differences in subjective ratings based on experience level or Replan Prompt Category. Statistics are shown in Table 7.

Table 7. Subjective Ratings Across Replan Prompt Categories

Metric	Replan Prompt Category	Mean	Std Dev	Mode	Median
Confidence self-rating	Low Replan Values	3.02	0.77	3	3
	Medium Replan Values	2.83	0.58	3	3
	High Replan Values	3.00	0.49	3	3
Workload self-rating	Low Replan Values	2.81	0.72	3	3
	Medium Replan Values	2.58	0.51	3	3
	High Replan Values	2.72	0.67	3	3

5. CONCLUSIONS

An experiment was conducted to examine the impact of allowing an operator to adjust the rate at which he or she is prompted to view new plans generated by an automated planner when supervising multiple heterogeneous unmanned vehicles. This capability was used in 62% of all trials and was used heavily when the initial replan rate was high (120 seconds between replan prompts).

Results showed that operators favored a particular replan interval, regardless of their initial replan rate. For the purposes of data analysis, operators were partitioned into groups based on whether they preferred low, medium, or high replan intervals. There was no significant difference in workload, performance, or self-ratings between these groups.

There was, however, a difference in performance based both on the initial replan interval and the average replan interval set through the Replan Dial. The results of this experiment confirmed the previous experiment by showing that rapid rates of replanning can cause lower performance. It was shown in this experiment, however, that based on increasing the average replan interval set through adjusting the Replan Dial, the time that targets were tracked increased, but the number of hostiles destroyed decreased. This is probably due to the fact that destroying a hostile was almost always given a higher priority than tracking a target. Upon faster replan methods the system will favor destroying newly discovered hostiles at the cost of tracking known targets, due to limited UxV resources.

Although differences in performance were found between experienced and inexperienced users of the testbed were found, these difference were not unexpected, as experienced operators would be expected to perform slightly better.

These results indicate that providing operators with the ability to adjust the rate at which they are prompted to view automation-generated plans did have an impact on overall mission performance and workload. Operators, on average, chose to decrease their workload when initially prompted at 30 or 45 second intervals, but chose to increase their workload when initially prompted at 120 second intervals. Significant differences in performance occurred based on both the initial replan interval and the average interval that operators chose. Further research is necessary to determine the impact of this method, but it is clear that future unmanned vehicles systems designs should incorporate the flexibility to allow operators to adjust the frequency at which the automation generates new plans for approval.

ACKNOWLEDGMENTS

This research is funded by the Office of Naval Research (ONR) and Aurora Flight Sciences. Olivier Toupet of Aurora Flight Sciences provided extensive support to generate the algorithms that run the simulation. Professor John How and the members of the MIT Aerospace Controls Laboratory provided the experimentation testbed support.

REFERENCES

Clare, A.S., Hart, C.S. & Cummings, M.L. (2010). Assessing operator workload and performance in expeditionary multiple unmanned vehicle control. *48th AIAA Aerospace Sciences Meeting,* Orlando, FL.

Cummings, M.L., Bruni, S., Mercier, S. & Mitchell, P.J. (2007). Automation architecture for single operator, multiple UAV command and control. *The International Command and Control Journal,* 1 (2)**,** 1-24.

Cummings, M.L., Clare, A.S. & Hart, C.S. (2010). The role of human-automation consensus in multiple unmanned vehicle scheduling. *Human Factors*, in press.

Cummings, M.L. & Guerlain, S. (2007). Developing operator capacity estimates for supervisory control of autonomous vehicles. *Human Factors,* 49 (1)**,** 1-15.

Cummings, M.L. & Nehme, C.E. (2009). Modeling the impact of workload in network centric supervisory control settings. *2nd Annual Sustaining Performance Under Stress Symposium.* College Park, MD.

Fisher, S. (2008). *Replan understanding for heterogenous unmanned vehicle teams*, Masters of Engineering thesis, Massachusetts Institute of Technology, Cambridge, MA.

Naval Studies Board (2005). *Autonomous vehicles in support of naval operations,* Washington DC, National Research Council.

Smith, P.J., McCoy, E. & Layton, C. (1997). Brittleness in the design of cooperative problem-solving systems: The effects on user performance. *IEEE Transactions on Systems, Man, and Cybernetics*, 27 (3)**,** 360-370.