# The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling

M. L. Cummings, Andrew Clare, and Christin Hart

Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding Author:
M.L. Cummings
Massachusetts Institute of Technology
77 Massachusetts Avenue 33-311
Cambridge, MA 02139
(617) 252-1512
(617) 253-4196 (Fax)
missyc@mit.edu

**ABSTRACT**

**Objective:** This study examined the impact of increasing automation replanning rates on operator performance and workload when supervising a decentralized network of heterogeneous unmanned vehicles. **Background:** Futuristic unmanned vehicles systems will invert the operator-to-vehicle ratio so that one operator can control multiple dissimilar vehicles, connected through a decentralized network. Significant human-automation collaboration will be needed due to automation brittleness, but such collaboration could cause high workload. **Method:** Three increasing levels 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. **Results:** Rapid replanning can cause high operator workload, ultimately resulting in poorer overall 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, as well as negative attitudes towards unmanned aerial vehicles in general. Participants with video game experience tended to collaborate more with the automation, which resulted in better performance. **Conclusion:** In decentralized unmanned vehicle networks, operators who ignore the automation's requests for new plan consideration and impose rapid replans both increase their own workload and reduce the ability of the vehicle network to operate at its maximum capacity. **Application:** These findings have implications for personnel selection and training for futuristic systems involving human collaboration with decentralized algorithms embedded in networks of autonomous systems.

**KEYWORDS:** multiple unmanned vehicles, human supervisory control, workload, human-automation collaboration, scheduling, vehicle routing, trust, human-automation consensus, operator-to-vehicle ratio, decentralized network, automation brittleness, task allocation, unmanned aerial vehicles

## INTRODUCTION

In current unmanned vehicle systems (UxVs), particularly those that include unmanned aerial vehicles (UAVs), more than one human operator is required for control and supervision. However, futuristic systems will invert the operator-to-UxV ratio so that one operator can control multiple UxVs (Cummings, Bruni, Mercier, & Mitchell, 2007; Franke, Zaychik, Spura, & Alves, 2005). To accomplish this future vision, significant collaborative autonomy must be embedded within and across teams of unmanned vehicles so that they can execute basic operational and navigation tasks with little human input. This architecture will require automated planners, which are faster than humans at path planning and resource allocation in multivariate, dynamic, time-pressured environments. However, human management of the automated planners will be critical, as automated planners cannot always generate accurate solutions in the presence of unknown variables. Though fast and able to handle complex computation far better than humans, computer optimization algorithms are notoriously "brittle" in that they can only take into account those quantifiable variables identified during the design stages as critical (Silverman, 1992; Smith, McCoy, & Layton, 1997).

Networked UxV operations can take the form of decentralized architectures, i.e., because of nearly constant sharing of information between and across network agents, no one vehicle is a central point of planning. In such decentralized networks, each vehicle computes its best plan with locally shared information, and no globally optimal plan exists since each vehicle strives to maintain the best plan with possibly limited information. Although the decentralized approach is superior to centralized in terms of protection against network vulnerabilities caused by bandwidth limitations and reliance on specific vehicles for critical tasks, it presents human supervisory control difficulties as operators who attempt to locally control a vehicle will not likely have an understanding of how this action could affect the network, since behavior is emergent.

Another significant concern in this concept of one operator supervising multiple collaborative UxVs involves the potential for high workload. Due to the dynamic and uncertain nature of the environment, control of decentralized UxVs requires rapid automated schedule replanning. Whereas embedded algorithms can typically generate new schedules in just seconds, often these schedules are not truly optimal, i.e., the optimized *a priori* cost function may not reflect the true state of all variables in dynamic scenarios. Moreover, automation will consider even a solution that is 0.1% better (given its hard coded objective function) as a better solution.

In command and control scenarios, the notion of "better" can be very different between automated and human decision makers. Since typically such algorithms do not account for satisficing behavior (Simon, et al., 1986) and cannot consider all variables (especially qualitative one), ultimately humans may be able to generate "better" solutions (in their view) than the automation. An important, but rarely addressed question in human supervision of collaborative networks of UxVs involves the impact of the rate of new plan generation by the automation that requires human approval, i.e., how often should the human be tasked to approve and possibly modify new UxV schedules that the automation deems to be better?

Operators engaged in these dynamic, high workload environments must both concentrate attention on the primary task (i.e., monitoring vehicle progress and identifying targets) and also be prepared for automation replan alerts. This need to concentrate on a task, yet maintain a level of attention for alerts requires both interrupt and task-driven processing. The allocation of attention between these two incurs cognitive costs that negatively impact overall system performance (Miyata & Norman, 1986). Poor attention allocation has been shown to be a significant contributor to poor operator performance in single operator control of multiple unmanned vehicles (Crandall & Cummings, 2007; Goodrich, Quigley, & Cosenzo, 2005).

In order to address these issues, given a simulation of a network of decentralized collaborative UxVs (Valenti, Bethke, Fiore, How, & Feron, 2006) supervised by a single operator under time pressure, an experiment was conducted to determine how increasing automation replanning rates in terms of new plan generation requiring approval affects operator and system performance. The assumed mission is one of search, track, and destruction of enemy targets, a typical command and control mission.

## METHOD

### Apparatus

This experiment employed a collaborative multiple UxV simulation environment that leverages decentralized algorithms for vehicle routing and task allocation. This simulation environment functions as a computer simulation but also supports actual flight and ground capabilities (Valenti, et al., 2006); all the decision support displays discussed in this paper have operated actual small air and ground UxVs.

Participants were responsible for one rotary-wing UAV, one fixed wing UAV, one Unmanned Surface Vehicle (USV), and a Weaponized UAV (WUAV). The UAVs and USV were responsible for searching for targets. Once a target was found, the operator was alerted to perform a target identification task (i.e., hostile, unknown, or friendly), along with assigning an associated priority level (i.e., high, medium, low). Then, hostile targets were tracked by one or more of the vehicles until the human operator approved WUAV missile launches. A primary assumption was that operators had minimal time to interact with the displays due to other mission-related tasks.

Operators had two exclusive tasks that could not be performed by automation: target identification and approval of all WUAV weapon launches. Operators also created search tasks, which dictated on the map those areas the operator wanted the UxVs to specifically search. Operators also had scheduling tasks, but these were performed in collaboration with the automation; when the autonomous planner recommended schedules, operators accepted, rejected, or modified these plans.

The autonomous planner only communicated to the vehicles a prioritized task list, and the vehicles sorted out the actual assignments among themselves by "bidding" on the tasks they felt they could accomplish, which was manifested through a consensus-based auction algorithm (Choi, Brunet, & How, 2009). In the course of this market-based auction scheme, vehicles bid on tasks while attempting to minimize the revisit times between tasks.

Operators were shown the results of this bidding process through the display that showed unassigned tasks that could not be completed by one or more of the vehicles. However, if unhappy with the UxV-determined search or track patterns, operators could create new tasks, in effect forcing the decentralized algorithms to reallocate the tasks across the UxVs. This human-automation interaction scheme is one of high level *goal-based* control, as opposed to more low-level *vehicle-based* control. Operators could never directly individually task a single vehicle.

Participants interacted with the simulation via two displays. The primary interface is a map display (Figure 1). The map shows both geo-spatial and temporal mission information (i.e., a timeline of mission significant events), and supports an instant messaging "chat" communication tool, which provides high level direction and intelligence. Icons represent vehicles, targets of all types, and search tasks, and the symbology is consistent with MIL-STD 2525 (DoD, 1999).

In this interface, operators identify targets, approve weapons launches, and insert new search tasks as desired or dictated via the chat box. The performance plot in Figure 1 gives operators insight into the automated planner performance, as the graph shows expected (red) versus observed (blue) performance of the system. Performance for this graphical representation is defined as a priority-level weighted sum of the tasks assigned in each plan with decrements to the performance score based on deviations from the desired arrival times to each task, i.e., the score decreases more when vehicles are late in task execution. When the expected performance score is above the observed score, the autonomous planner is effectively proposing that better performance could be achieved if the operator accepts the proposed plan (based on the planner's prediction of how the vehicles will bid on the tasks). When the observed performance curve surpasses the expected curve, the human operator has changed the tasking, which causes better than expected automation performance.

For this experiment, the automation's *a priori* cost function was to track targets in order of their respective priorities and minimize switching times between vehicles. This cost function, as well as bidding iterations between the vehicles, was not explicitly shown to operators, and operators were told that their objectives were to search the largest area possible, track all found targets, and destroy all known hostile targets. Participants were told that there could be a discrepancy between their objectives and those of the automation since this is a realistic brittleness problem as previously described.

The decentralized planning automation can generate a new "improved" plan approximately every second based on the static cost function, but these improvements can be quite small, e.g., ~.1%, assuming something changed in the environment since the last plan. However, when to show a new plan to the human is the subject of the experiment that will be discussed in the next section since a 1s replanning interval is not feasible for operators. When the automation generates a new plan that is “better” than the last, the Replan button turns green and flashes, and when selected, the operator is taken to the Schedule Comparison Tool (SCT), detailed in the next section.

Operators can elect to select the Replan button at anytime, even when it is not lit. The SCT display then appears, showing three geometrical forms colored gray, blue, and green at the top of the display (Figure 2), which are configural displays that enable quick comparison of schedules. The left form (gray) is the current UxV schedule. The right form (green) is the latest automation-proposed schedule. The middle working schedule (blue) is the schedule that results from user plan modification. The rectangular grid on the upper half of each shape represents the estimated area of the map that the UxVs will search according to the proposed plan. The hierarchical priority ladders show the percentage of tasks assigned in high, medium, and low priority levels.

When the operator first enters the SCT, the working schedule is identical to the proposed schedule. The operator can conduct a “what if” query process by dragging the desired unassigned tasks into the large center triangle. This query forces the automation to generate a new plan if possible, which becomes the working schedule. The configural display of the working schedule alters to reflect these changes. However, due to resource shortages, not all tasks can be assigned to the UxVs, representative of real world constraints. The working schedule configural display updates with every individual query so that the operator can leverage direct-perception interaction (Gibson, 1979), to quickly compare the three schedules.

This “what-if” query, which essentially is a preview display (Wickens & Hollands, 2000), represents a collaborative effort between the human and automation (Layton, Smith, &

McCoy, 1994). Operators adjust team coordination at the task level as opposed to individual vehicle tasking, which has been shown to improve single operator control of a small number of multiple independent robots (Goodrich, et al., 2007). Details of the interface design and usability testing can be found in Fisher (2008).

**Participants and Procedure**

Thirty-one participants volunteered for this study (twenty-four men and seven women). Ages ranged from 19 to 44 years with a mean of 23.8 years and standard deviation (sd) of 4.8 years. Sixty-eight percent of participants had military experience (ROTC, military academy, or active duty). Each participant filled out a demographic survey prior to the experiment that included age, gender, occupation, military experience, video gaming experience, and perception of UAVs.

The experiment was conducted on a Dell Optiplex GX280 with a Pentium 4 processor and an Appian Jeronimo Pro 4-Port graphics card. The display's resolution was 1280x1024 pixels. A self-paced, slide-based tutorial was provided that explained the interfaces, as well as a practice session. Training took approximately thirty minutes, followed by three counterbalanced ten-minute test sessions, representing three possible replan intervals (detailed below). Each scenario was different, but similar in difficulty.

**Experimental Design**

In order to determine the specific replanning intervals that would result in high operator workload, a previously–developed single operator, multiple UxV discrete event simulation (DES) model was used (Cummings & Nehme, 2010). Given specific capabilities of different unmanned vehicles (i.e., autonomy expressed through neglect times), human operators (i.e., task error rates, attention allocation across tasks), and the interface (i.e., expected task response times), this model predicts operator utilization, also known as percent busy time. Utilization represents the time operators are required to interact with a system divided by the total time of possible interaction. Monitoring time is not included in this metric.

Prior human performance models using steady state queuing models estimate that humans cannot successfully perform at utilization rates > 70% (Rouse, 1983; Schmidt, 1978). Empirical studies based in single-operator supervisory control of multiple entities have shown this prediction to be a reliable predictor of high workload (Cummings & Guerlain, 2007; Cummings & Nehme, 2009). Vehicle performance data from the test bed and preliminary operator performance estimates were used to predict the rates of new automation-generated schedules requiring operator approval (called "replanning rates"). It was predicted that replanning rates of 120s, 45s, and 30s would produce utilization levels of 55%, 70%, and 85%, respectively. It was hypothesized that operators would perform satisfactorily at the 45s and 120s intervals, but the 30s interval would cause a significant drop in performance due to the increase of utilization over 70%.

For these fixed intervals, the embedded autonomy produced a plan that was, on average, 7-20% better than the previous plan, but there were no convergent values for each of the three replanning intervals, i.e., the 30s interval could produce a plan 20% better, and the 120s a 7% better plan, and vice versa. In fact, the automation was brittle in that it occasionally would recommend a plan worse that the current plan. In cases where nothing in the environment had changed, operators were instructed to create a search task anywhere of their choosing, which would then cause the algorithms to compute a new best plan.

Performance dependent variables included the percentage of area covered, number of targets found, number of hostile targets destroyed, and the average operator utilization for a given session. For utilization, operators were considered "busy" when performing one or more of the following tasks: creating search tasks, identifying and designating targets, approving weapons launches, interacting via the chat box, and replanning in the SCT. All interface interactions were via a mouse with the exception of the chat messages, which required keyboard input.

In addition to objective workload, secondary workload was measured via reaction times to text message information queries, as well as reaction times when instructed to create search tasks via the chat tool. Such embedded secondary tools have been previously shown to be effective indicators of workload (Cummings & Guerlain, 2004).

## RESULTS AND DISCUSSION

A repeated measures Analysis of Variance (ANOVA) model was used for parametric dependent variables ($\alpha = 0.05$). For dependent variables that did not meet ANOVA assumptions, non-parametric analyses were used.

### Performance Metrics

The three mission performance metrics showed the same trend in that the 30s replan interval resulted in lower performance than the 45s and 120s intervals (Figure 3). Specifically, the omnibus area coverage test was significant (Figure 3a, $F(2, 56) = 16.213$, $p < 0.001$), and Tukey pairwise comparisons showed significant differences between the 30s and 45s intervals ($p < 0.001$), and between the 30s and 120s intervals ($p = 0.001$). There was no significant difference between the 45s and 120s intervals for area coverage ($p = 0.120$). In terms of practical utility, those participants at the 45s and 120s replanning levels covered on average 10% more area than the 30s participants, which is significant in terms of actual military operations, e.g., additional missions do not have to be launched to cover this area.

As with the area coverage metric, the omnibus percentage of targets found was significant (Figure 3b, $F(2, 60) = 14.640$, $p < 0.001$). Tukey pairwise comparisons also showed that the 30s interval was different from 45s and 120s intervals ($p = 0.002$ and $p < 0.001$ respectively), but once again, the 45s and 120s intervals were not statistically different ($p = 0.096$).

For the percentage of hostile targets destroyed (Figure 3c), non-parametric tests were needed. The Friedman test showed a significant difference for Hostiles Destroyed ($\chi^2(2, N=31) = 9.152$, $p = 0.010$.) Mann-Whitney Dependent pairwise comparison tests showed the 30s interval was significantly different from the 45s interval ($p = 0.010$). Overall, the performance results showed that operators at the 30s replan interval consistently covered less area, found fewer targets, and destroyed fewer hostile targets as compared to the 45s and 120s intervals.

**Workload Metrics**

Utilization was significantly different across the three independent replan intervals ($F(2, 60) = 10.601$, $p < 0.001$). Tukey pairwise comparisons showed that the 30s interval was significantly different from the 45s and 120s intervals ($p = 0.002$ and $p < 0.001$, respectively). As with the performance results, the 45s and 120s intervals were not statistically different ($p = 0.354$). Overall, the utilization metric showed that operators had a higher workload at the 30s replan interval (Figure 4). Non-significant results were obtained for the secondary workload metrics (chat reaction time and prompted search task reaction time), as well as subjective workload measured on a 4-point Likert scale (from 'idle' to 'extremely busy').

## DISCOVERY OF UNDERLYING SUBPOPULATIONS

Although these initial results confirmed our original hypothesis that people would perform significantly worse under the 30s replanning interval than the 45s and 120s intervals, the validation of the previously-discussed queuing-based model's predictive capability led to a significantly more in-depth analysis. To this end, the discrete event model's service time and event generation distributions were updated with the actual observed distributions, a common practice in simulation modeling (Law & Kelton, 2000). However, fitting of the data was unexpectedly a difficult task, which led to discovery of underlying participant populations and new insights into the experiment.

When attempting to fit distributions to both service time and user-generated event arrival times, it became apparent that the data were multimodal. Further analysis revealed that a major source of the multimodal distribution problem was that in many cases, despite careful training, participants often chose to ignore the automated replanning reminders and elected to replan on their own schedule. For example, regardless of whether they were prompted to evaluate a new schedule in the SCT every 30s, 45s, or 120s, some participants simply ignored the alert and performed other tasks like communication or search task creation. These participants also entered the SCT at times when the green button was not illuminated and asked the automation for a new plan.

This behavioral trend led to partitioning the participants into three categories: 1) those participants with statistically significant different rates of replanning across the three intervals, which aligned with the automation's prompting, called Consenters, 2) those that had no statistically significant differences in the rates of replanning across the three trials, as they routinely ignored the automated alerts and chose to enter the SCT at other times, called

Dissenters, and lastly 3) Mixed Consenters, whose rates occasionally aligned with the automation's prompting, but who also used the SCT between the automation's recommendations. These three populations were fairly evenly split (29%, 39%, and 32%, respectively). Table 1 shows the intended replan rates and those rates adopted by participants in the three categories.

The label Consenter does not imply that participants agreed with the automation's new proposed plan, only that they agreed to consider it. Regardless of consent or dissent status, participants were not statistically different in their acceptance of the automation's proposed plans, in that all three groups modified the automation's suggestions ~33% of the time, which generally improved the system performance by ~50% in terms of targets found and area covered. However, operator changes in the schedule generally did not improve (or detract from) targets tracked or hostiles destroyed. The Dissenters under high workload (i.e., 30s replan interval) modified proposed plans slightly more, ~43% of the time. Even though under the greatest time pressure, the dissenters chose to manipulate the schedule the most often.

Table 1 reveals two interesting points. First is the observed replan rate for the Consenters at the 120s replan interval, which is well below the expected interval. Only two participants consistently waited the full two minutes for the automation's new proposed plan. Even though the Consenters generally followed the automation's suggestions for new plan consideration for the 30s and 45s replan intervals (which were counterbalanced), they generally could not wait the full 120s.

Of additional interest in Table 1 is the nearly identical self-imposed replanning rate of the Dissenters. On average, regardless of whether the automated replan alert signaled that a new plan was available, the Dissenters entered the SCT to query the automation for a new plan approximately every 30 seconds. As discussed previously, participants could query the automation for a "what-if" solution anytime. Dissenters elected to do this far more often than the other groups. However, there was no statistical difference between the Consenters and Dissenters in terms of how long they spent using the SCT once they actually decided to replan.

Given the three distinct sub-populations of Consenters, Dissenters, and Mixed Consenters, an additional analysis was conducted to determine how the degree of consent or dissent influenced performance. Using Spearman Rho correlations, increasing dissent led to increased utilization ($\rho = 0.331$, $p = 0.001$), as well as decreased area coverage ($\rho = -0.251$, $p = 0.017$). There were no significant correlations with other performance variables. These relationships are depicted in Figures 5a and 5b. Dissenters worked more than the other groups, but their performance declined in terms of area coverage. Despite their efforts, the Dissenters did not perform as well as those participants who collaborated with the autonomous planner in terms of reviewing (but not necessarily accepting) proposed schedule changes.

In an attempt to shed light on any possible associations that could elucidate this finding, several demographic factors were examined. Spearman Rho correlations showed no significant correlations for military experience (either the presence or amount of experience), nor for age in terms of consent/dissent status. Participants were fairly homogenous in these demographics, as most were in their middle twenties, and the majority had some form of military experience. One interesting correlation involved participants' views of UAVs in general. Prior to the experiment, participants were asked to give general impressions of UAVs, which were coded into a 5-point Likert scale (1 = intense dislike, 5 = really like). The Spearman Rho correlation between this response and the degree of consent was -0.234 ($p = 0.024$). Not surprisingly, dissenters weakly correlated with those who did not like UAVs.

One additional important correlation was the association between degree of consent and video game experience, which was moderate-strong ($\rho = -0.372$, $p < 0.001$). The Consenters and Mixed Consenters had the most amount of video game experience, on average reporting weekly to monthly gaming experience, whereas the Dissenters reported significantly less (none to monthly). Previous studies have shown that there are possibly neural bases for improved performance in video game environments, such as increased dopamine release (Koepp, et al., 1998) and increased activation and connectivity in the mesocorticolimbic system, particularly for males (Hoeft, Watson, Kesler, Bettinger, & Reiss, 2008). Additional research has shown that video game play can improve visual attentional processing in divided attention, multi-task environments similar to the one described here (Green & Bavelier, 2003). However, other research has shown that people who are heavy multi-taskers in multimedia environments do not perform as well as those who are not (Ophir, Nass, & Wagner, 2009).

It appears that the gamers in this study were able to self-regulate their pace so that they were able to effectively divide their attention. The Dissenters, however, elected to rapidly task themselves with marked regularity, and in doing so, performed significantly worse. Thus the gamers in this study seemed to have more patience and were willing to allow the automation to do its job, which ultimately led to better performance than those with little to no gaming experience.

As has been noted in several other related studies (Lee & See, 2004; Moray, Inagaki, & Itoh, 2000), trust is a critical component of any supervisory control system with embedded automation. Although some component of trust is no doubt a factor in this study given the weak correlation between dislike of UAVs and Dissenters, from a performance perspective, the Consenters did not trust the system any more than the Dissenters. Whereas the Consenters and

Dissenters each modified the automated plans about 33% of the time, the Consenters simply agreed to the *pace* of replanning and were not task switching as frequently as the Dissenters.

These findings have implications for personnel selection and training for futuristic systems involving human collaboration with networks of unmanned systems with embedded autonomous planners. Although a positive association was found between performance and video game experience in this study, causality is still unclear. Gamers could be an attractive population for future unmanned vehicle operators, but recent research has also shown that people can improve their visual attentional processing after just 10 days of playing video games (Green & Bavelier, 2003). Thus, the cognitive benefits of video gaming could be gained through training.

In addition, the possible influence of participants' attitudes towards unmanned systems cannot be ignored. Previous research has shown that there are cultural barriers to acceptance of unmanned vehicles in some military populations (Cummings, 2008; Cummings & Morales, 2005). This present study indicates that such a lack of acceptance not only has cultural implications, but performance ones as well.

## CONCLUSIONS

An experiment was conducted to investigate the impact of automated planning rates for single operator control of multiple unmanned vehicles in a decentralized network. Participants performed significantly better when the automation prompted them to replan every 45s and 120s, as compared to replanning every 30s. In determining probability distributions to validate a discrete event operator workload model based on this experimental data, three distinct sub-groups were unexpectedly discovered, almost evenly split across the data: (1) Consenters who generally followed the automation's prompting for when to consider new plans (29%), (2) Dissenters who almost always ignored the automation's prompts (39%), and lastly (3) Mixed Consenters whose behaviors were split between these two extremes (32%). A lack of replanning rate consensus between the automation and operators was significantly correlated with degraded performance and a dislike of UAVs in general. In contrast, higher degrees of consent were associated with better performance as well as video game experience.

Networks of unmanned vehicles will most likely be operated in the future with similar decentralized planners since such planners are not as vulnerable as centralized planners and can offer performance guarantees such as conflict-free task assignment (Alighanbari & How, 2006). This significant level of autonomy onboard each UxV will cause operators to supervise the system from a goal-based perspective as opposed to more vehicle-based control seen in today's UAV systems. Because of the nearly-constant replanning that must occur on the part of the decentralized UxV planners, operators will have to adapt a more collaborative approach to mission supervision. As shown in this study, operators who ignore the automation's requests for new plan consideration and self-impose rapid replans both increase their own workload and reduce the ability of the network of vehicles to operate at its maximum capacity.

This study raises several important areas of future research in unmanned vehicle supervision, such as the need for human-computer collaboration in decentralized automated planning environments, the influence of video game experience in such collaborative environments, the negative impact of rapid multi-tasking, and the role underlying attitudes have on overall system performance. These questions are not just critical for unmanned vehicle systems, but for all supervisory control systems, especially as they grow more complex with increasing embedded autonomy.

## ACKNOWLEDGMENTS

This research is sponsored by Aurora Flight Sciences and the Office of Naval Research. Olivier Toupet of Aurora Flight Sciences provided extensive algorithm support. Ian Davies and Pierre Maere assisted in data analysis, and Professor John How and the MIT Aerospace Controls Laboratory provided test bed support.

## REFERENCES

Alighanbari, M., & How, J. P. (August 2006). *Robust decentralized task assignment for cooperative UAVs.* Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit, Keystone, CO.

Choi, H. L., Brunet, L., & How, J. P. (2009). Consensus-based decentralized auctions for robust task allocation. *IEEE Transactions on Robotics*, *25*(4), 912–926.

Crandall, J. W., & Cummings, M. L. (May 2007). *Attention allocation efficiency in human-uv teams*. Proceedings of the AIAA Infotech@Aerospace Conference, Rohnert Park, CA.

Cummings, M. L. (2008). Of shadows and white scarves. *C4ISR Journal*, *7*, 17–18. Retrieved from http://www.c4isrjournal.com/story.php?F=3581309

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

Cummings, M. L., & Guerlain, S. (2004). Using a chat interface as an embedded secondary tasking tool. In D. A. Vincenzi, M. Mouloua, & P. A. Hancock (Eds.), *Human performance, situation awareness and automation: Current research and trends. HPSAA II* (Vol. 1, pp. 240-248). London, UK: Routledge.

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

Cummings, M. L., & Morales, D. (2005). UAVs as tactical wingmen: Control methods and pilots' perceptions. *Unmanned Systems, 23*(1), 25–27.

Cummings, M. L., & Nehme, C. E. (2010). *Modeling the impact of workload in network centric supervisory control settings*. In S. Kornguth, R. Steinberg, & M. D. Matthews (Eds.), *Neurocognitive and physiological factors during high-tempo operations* (pp. 23-40). Surrey, UK: Ashgate.

Fisher, C. S. (2008). *Real-time, multi-agent system graphical user interface.* Cambridge, MA: MIT.

Franke, J., Zaychik, V., Spura, T., & Alves, E. (June 2005). *Inverting the operator/vehicle ratio: Approaches to next generation UAV command and control.* Paper presented at the Association for Unmanned Vehicle Systems International and Flight International, Unmanned Systems North America, Baltimore, MD. Retrieved from http://www.atl.lmco.com/papers/1261.pdf

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Goodrich, M. A., McLain, T. W., Anderson, J. D., Sun, J., & Crandall, J. W. (2007). Managing autonomy in robot teams: Observations from four experiments. In *ACM/IEEE International Conference on Human-Robot Interaction* (pp. 25–32). New York, NY: Association for Computing Machinery.

Goodrich, M. A. Quigley, M., & Cosenzo, K. (2005). Task switching and multi-robot teams. In L. E. Parker, F. E. Schneider, & A. C. Schultz (Eds.), *Multi-robot systems: From swarms to intelligent automata* (Vol. 3, pp. 185-195). Dordrecht, Netherlands: Springer.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534–537.

Hoeft, F., Watson, C. L., Kesler, S. R., Bettinger, K. E., & Reiss, A. L. (2008). Gender differences in the mesocorticolimbic system during computer game-play. *Journal of Psychiatric Research*, *42*, 253–258.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., ... Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, *393*, 266–268.

Law, A. M., & Kelton, D. (2000). *Simulation modeling and analysis* (3rd ed.): Boston, MA: McGraw-Hill.

Layton, C., Smith, P. J., & McCoy, C. E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. *Human Factors*, *36*, 94–116.

Lee, J. D., & See, K. A. (2004). Trust in technology: Designing for appropriate reliance. *Human Factors*, *46*, 50–80.

Miyata, Y., & Norman, D. A. (1986). Psychological issues in support of multiple activities. In D. A. Norman & S. W. Draper (Eds.), *User centered system design: New perspectives on human-computer interaction* (pp. 265–284). Hillsdale, NJ: Erlbaum.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, *6*(1), 44–58.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences, 106*(37), 15583–15587.

Rouse, W. B. (1983). *Systems engineering models of human machine interaction*. New York, NY: North Holland.

Schmidt, D. K. (1978). A queuing analysis of the air traffic controller's workload. *IEEE Transactions on Systems, Man, and Cybernetics*, *8*(6), 492–498.

Silverman, B. G. (1992). Human-computer collaboration. *Human-Computer Interaction*, *7*(2), 165–196. Simon, H. A., Dantzig, G. B., Hogarth, R., Piott, C. R., Raiffa, H., Schelling, T. C., . . . Winter, S. (1986). Decision making and Problem solving. In *Research briefings 1986: National Academy of Sciences report of the Research Briefing Panel on Decision Making and Problem Solving* (pp. 17–36). Washington, DC: National Academy Press.

Smith, P., 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-Part A: Systems and Humans*, *27*, 360–371.

U.S. Department of Defense. (1999). *Interface standard, common warfighting symbology MIL-STD-2525B* (MIL-STD-2525B). Washington, DC.

Valenti, M., Bethke, B., Fiore, G., How, J. P., & Feron, E. (August 2006). *Indoor multi-vehicle flight testbed for fault detection, isolation, and recovery*. Proceedings of the AIAA Guidance, Navigation and Control Conference and Exhibit, Keystone, CO.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

**Table 1: Consensus Status and Replan Intervals**

| | 30s Replan Interval | 45s Replan Interval | 120s Replan Interval |
|---|---|---|---|
| **Dissenters** | 32 ± 6.9s | 37 ± 13.6s | 38 ± 11.7s |
| **Mixed Consenters** | 35 ± 6.5s | 36 ± 8.8s | 65 ± 26.1s |
| **Consenters** | 33 ± 6.3s | 46 ± 6.3s | 79 ± 18.7s |

**LIST OF FIGURES**

Figure 1: Map Display

Figure 2: The Schedule Comparison Tool

Figure 3: Performance Metric Comparison
(a) Area Coverage
(b) Targets Found
(c) Hostiles Destroyed

Figure 4: Utilization Results

Figure 5: The Impact of Consensus on Performance
(a) Utilization
(b) Area Coverage

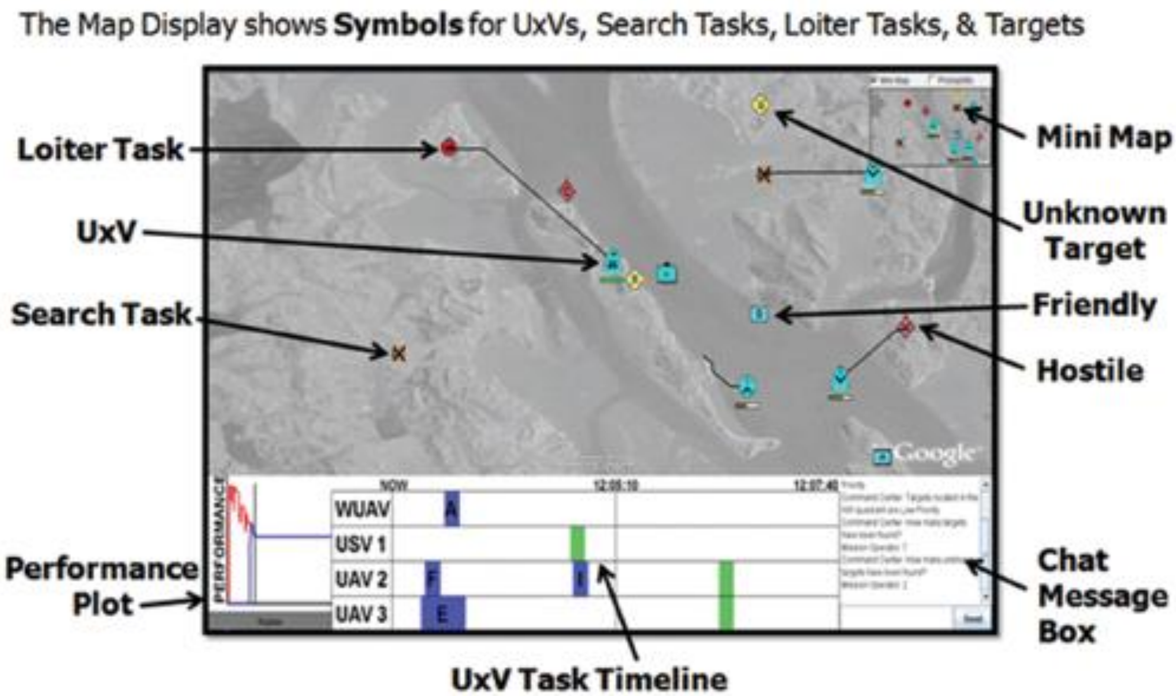


*Figure 1.* Map display. UxV = unmanned vehicle system.

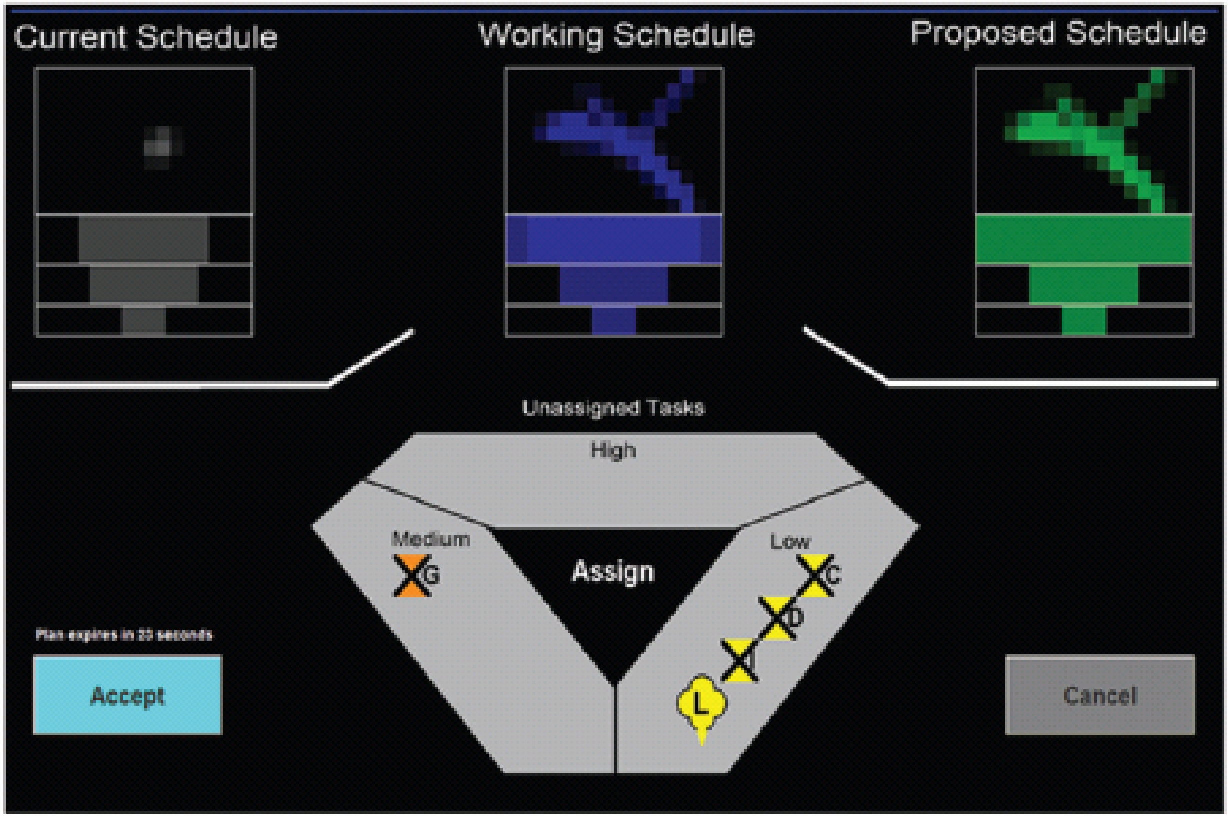


*Figure 2*. The schedule comparison tool.

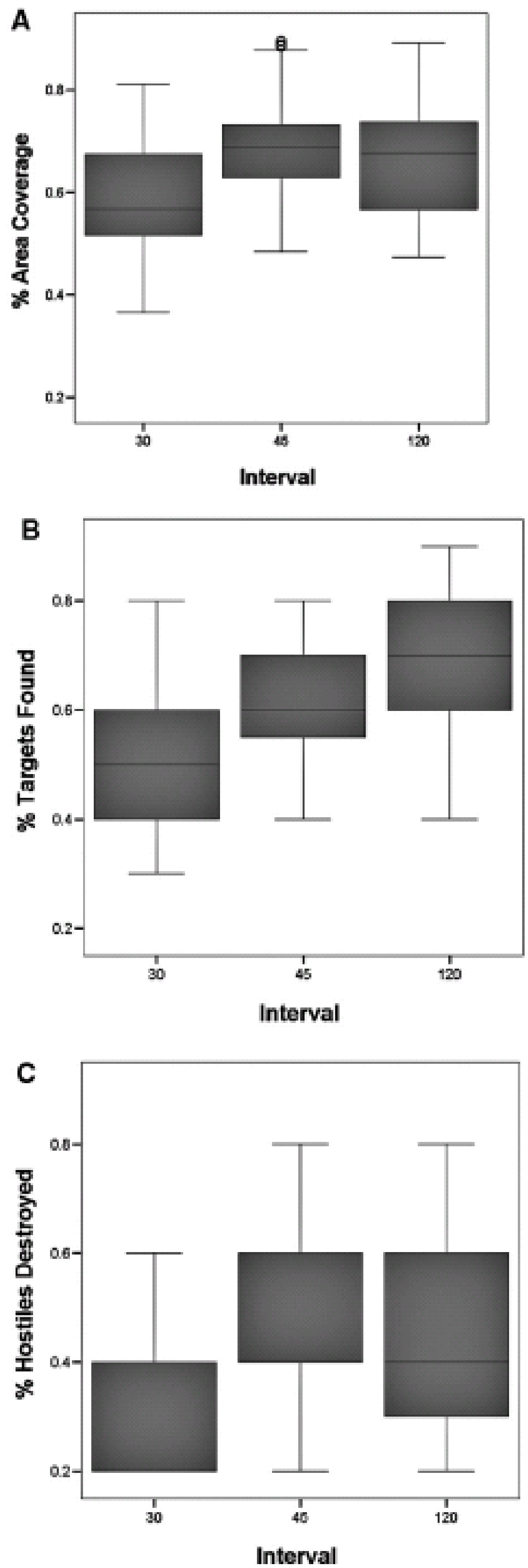


*Figure 3.* Performance metric comparison: (A) area coverage, (B) targets found, (C) hostiles destroyed.

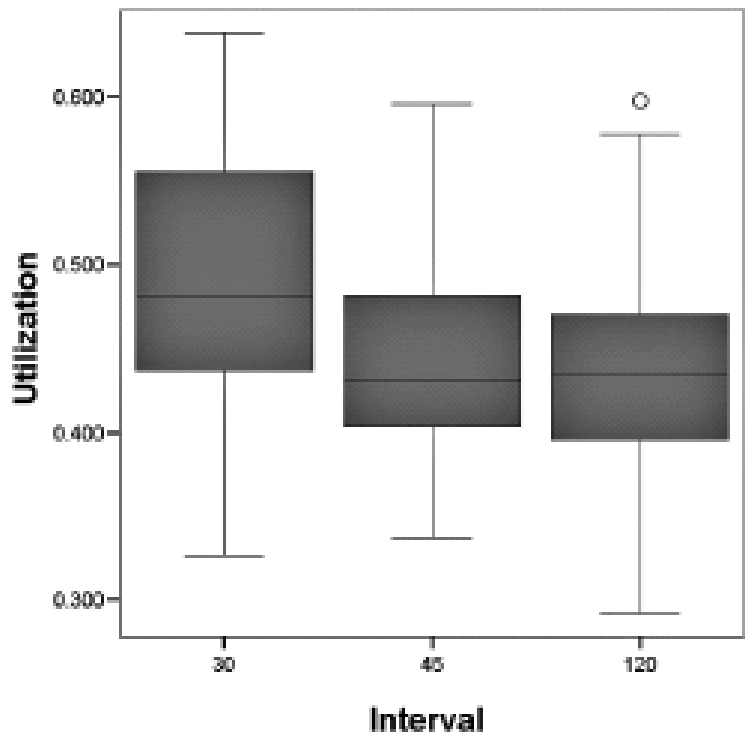


*Figure 4.* Utilization results.

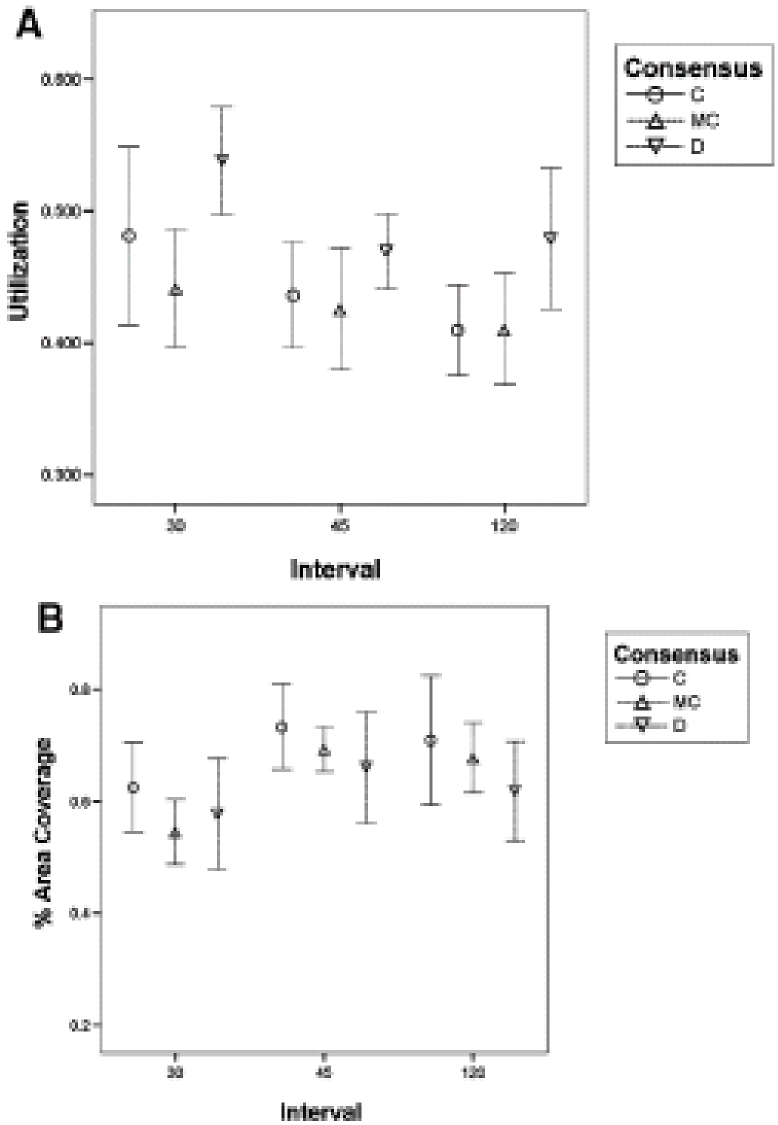


*Figure 5.* The impact of consensus on performance: (A) utilization, (B) area coverage. C = consenters; MC = mixed consenters; D = dissenters.

## BIOGRAPHIES

**MARY L. CUMMINGS** received her Ph.D. in Systems Engineering from the University of Virginia in 2003. She is currently an associate professor in the MIT Aeronautics & Astronautics Department, and is the director of the Humans and Automation Laboratory.

**ANDREW CLARE** is a Ph.D Candidate in the Humans and Automation Laboratory at the Massachusetts Institute of Technology, Cambridge, MA.  He received his B.S. in Aerospace Engineering with Information Technology at the Massachusetts Institute of Technology in 2008.

**CHRISTIN HART** is pursuing her Master's degree in Aeronautical Engineering at the Massachusetts Institute of Technology, Cambridge, MA.  She received her B.S. in Aerospace Engineering from the U.S. Air Force Academy in 2008.