Predicting the Impact of Heterogeneity on Unmanned-Vehicle Team Performance

Carl E. Nehme[1], Ryan M. Kilgore[2], M. L. Cummings[1]
[1]*Humans and Automation Laboratory, Massachusetts Institute of Technology, Cambridge, MA*
[2]*Charles River Analytics, Inc., Cambridge, MA*

Several recent studies have addressed the possible impact of using highly autonomous platforms to invert today's multiple-operators-per-single-unmanned-vehicle control paradigm. These studies, however, have generally focused on homogeneous vehicle teams and have not addressed the potential effects of vehicle, capability, or mission type heterogeneity on operator control capacity. Important implications of heterogeneous unmanned teams include increases in the diversity of potential team configurations, as well as the diversity of possible attention allocation strategies that may be utilized by operators in managing a given vehicle team. This paper presents preliminary findings from a modeling and simulation effort exploring the impact of heterogeneity on the supervisory control of unmanned vehicle teams. Results from a discrete event simulation study suggest that performance costs of team heterogeneity are highly dependent on resultant changes in operator utilization. Heterogeneous teams that result in lower overall operator utilization may lead to improved performance under certain operator control strategies.

INTRODUCTION

More effective decision-support tools and increasing use of automation in unmanned vehicle systems has shifted the level of the human operator's responsibility from manual to supervisory control. At the supervisory control level, implementation details of higher-order mission tasking initiated by the human is delegated to vehicles' onboard automation (Sheridan, 1992). The potential workload reduction afforded by supervisory control systems can increase operator idle time. This spare time may be used as a force multiplier, allowing operators to supervise multiple vehicles simultaneously, thus inverting the currently required many-to-one ratio of operators to vehicles. Inverting this ratio may also allow for reduced manning, which is a current military operational concern.

To date, there has been a growing body of literature that has examined the capacity of single operators to supervise multiple unmanned vehicles (Cummings et al., 2007; Olsen & Wood, 2003; Ruff, Narayanan, & Draper, 2002). This research has mainly focused on the supervision of homogeneous unmanned vehicle (UV) teams, where all vehicles have identical operating characteristics, capabilities, and tasks. However, as the mission goals of unmanned systems become increasingly demanding, UV teams are likely to become more heterogeneous in nature.

Multiple dimensions of heterogeneity, including vehicle type, platform capabilities, and mission tasks, introduce a number of problems in applying previous models of homogeneous UV team performance to heterogeneous systems. Analysis of the potential vehicle/capability/task permutations of large unmanned teams presents a complex and mathematically intractable problem. Moreover, the method by which operators attend to and interact with individual vehicles is likely to be affected by team heterogeneity, with significant impact on operator workload and overall system performance. A better general understanding of not only the various operator strategies for supervising unmanned vehicle teams, but how these strategies contribute to system performance under varying types of team heterogeneity, is needed. This paper describes a preliminary effort to address this need through a modeling and simulation-based exploration of the impact of heterogeneity on the supervisory control of unmanned vehicle teams.

MODELING INTERACTION WITH HETEROGENEOUS UNMANNED SYSTEMS

Previous research by Olsen and Goodrich (2003) and Crandall et al. (2005) introduced several important temporal-based metrics for describing operator interaction with multiple robots. *Neglect time* (NT) was defined as the expected amount of time that a robot (which is representative of any unmanned vehicle) can be ignored before its performance drops below some acceptable threshold. *Interaction time* (IT) was defined as the average time it takes for a human to interact with the robot to ensure it is still working effectively towards mission accomplishment.

Using these metrics as a starting point, we developed a discrete-event simulation (DES) model to examine the impact of operator attention allocation strategies on overall system performance when supervising heterogeneous unmanned vehicle teams. This DES model includes: (a) a queuing model describing a resource-limited human operator's serial interaction with multiple UVs; and (b) a component for the ability to measure overall system performance. Both components of this model are described briefly below. More complete and detailed descriptions of this model have been presented in previous publications (Cummings et al., 2007; Nehme & Cummings, 2007; Nehme et al., in submission).

The operator model is based on a single server queue with multiple input streams (one stream for each vehicle). Events in this queuing system represent operator tasks, such as vehicle path planning and health and status intervention. The arrival rate of events on a stream *i* is a function of the rate at which vehicle *i* needs operator attention as well as the level of operator situational awareness. Degraded situational awareness leads to a reduced arrival rate which represents operators taking longer to notice vehicle requests for attention. The service rate of events arriving from stream *i* is a function of the length time for which vehicle *i* needs service as well as any time penalties due to operator context switching. In the model, the effect of switching times is to increase the length of operator interaction times, an effect which has been captured in previous research on multi-vehicle control (Goodrich, Quigley, & Cosenzo, 2005). For this DES model, context switching was accounted for whenever the current vehicle's capability or its task type differed from that of the last vehicle serviced. By varying parameters such as the arrival rates of events to the system and the corresponding service rates of operator-vehicle interaction, a variety of UV systems can be rapidly modeled.

In addition, the model can capture different operator attention-allocation strategies. This is possible by varying the rate of additional operator-induced events which can represent different levels of aggressiveness in operator re-planning and through alternate operator preference ordering for servicing the different events (by varying the queuing discipline).

The second significant part of the model is its ability to capture several performance-related metrics, including operator utilization, average vehicle wait times, and overall mission success score. This is possible through the queue structure, which supports recording different event wait times, as well as the times for which the event server (the model of the human operator) is busy. The model's ability to capture performance-based metrics is important as it allows for the evaluation of the effectiveness of different team configurations based on their performance. Testing for system robustness can also be undertaken by varying input parameters such as the rate at which vehicles generate events (a representation of environmental unpredictability) and observing the effects on one more performance-based metrics.

In addition, by varying parameters such as the operator situational awareness model (Nehme, Crandall, & Cummings, 2008) or the operator attention-allocation strategies, the effect of modifying interface/decision support to encourage different operator behavior can be studied. However, interface effectiveness was not manipulated as an independent variable in this study.

EXPERIMENTAL STUDY

The focus of this experiment was to determine the effects of a subset of operator resource allocation strategies on mission performance for varying levels of heterogeneity in the unmanned vehicle team. Three independent variables were of interest in this experiment: team-heterogeneity, operator level of neglect strategy, and operator attention switching strategy, each described in detail below.

Table 1. Vehicle Team Configurations

	Vehicle1		Vehicle2		Vehicle3	
	NT	IT	NT	IT	NT	IT
Team1	60	10	60	10	60	10
Team2	30	10	60	10	60	10
Team3	60	10	60	10	120	10
Team4	30	10	60	10	120	10

Vehicle Team Heterogeneity

Four distinct levels of team-heterogeneity were examined in this experiment: one representing a homogeneous team, and the other three representing different heterogeneous team configurations. Vehicles differed across these team configurations with respect to their expected *neglect times*.

The first level, *Team1*, consisted of three identical UAVs performing an identical surface imagery task. The NT for each vehicle was drawn from a normal probability distribution with a mean of 60s and a standard deviation of 6s, or 10% of the mean.

The second level, *Team2*, was created by replacing a single UAV from the previous team with an unmanned surface vehicle (USV), also assigned a surface imagery task. The mean NT distribution for this new USV was 30s, with a standard deviation of 3s (again, 10% of the mean). This smaller NT indicates that this vehicle requires more frequent attention from the human operator (this may be due, for example, to a lower level of vehicle autonomy).

The third team, *Team3*, was created by replacing the task of a single UAV in the homogeneous team with a communications task. The means of the NT distributions for the UAVs performing the surface imagery task were 60s, whereas the mean of the distribution for the UAV performing the communications task was 120s (this may be due, for example, to the relative simplicity of communications relay tasking). Again, the standard distribution for each NT was 10% of the distribution mean.

Finally, *Team4* was created by having three different vehicle/task pairs, each having a different mean for their NT distributions. *Team4* consisted of a USV and a UAV each assigned a surface imagery task, as well as a UAV assigned a communications task. This level represents the maximum heterogeneity level.

For all vehicle types, the same function was used to generate the IT that the vehicle event would require (this is excluding an effects due to context switching). When vehicles did not have to wait past their neglect period for operator interaction (zero wait time), IT was selected from a distribution with a mean of 10s, and with a standard deviation of 3s. When an operator attended to a vehicle sometime after NT had expired, a penalty was added that was proportional to the wait time. This may be due for example to a vehicle spending additional fuel while loitering which would result in the operator having to re-plan the rest of the vehicle's path in order to meet the new fuel constraints. NT and IT data for each team configuration is summarized in Table 1.

Level of Neglect

Level of neglect (LON) represents the extent to which an operator attempts to manage individual automated vehicles. For example, a high LON would reflect a macro-management strategy where the operator attempts to service vehicle *i* at a frequency equivalent to NT_i, the rate at which vehicle *i* actually needs individual attention. A micro-management strategy on the other hand, is represented by a low LON, which means the operator services the vehicle more often than the vehicle needs attention.

For this experiment, the operator LON factor consisted of three levels*;* *Macro, Macro/Micro,* and *Micro*, representing alternate operator neglect strategies. The *Macro* neglect strategy represented a macro-management situation where the actual time a vehicle is neglected by the operator was equivalent to NT, excluding any effects due to loss of situational awareness or queuing wait times. The *Macro/Micro* strategy represents a medium level of neglect, where the rate of intended operator interaction was 0.75*NT. In this case, the operator is partly attempting to micromanage the vehicles, but doing so at a moderate level. Finally, the *Micro* strategy represented an extreme case of micromanaging where the rate of intended interaction was equivalent to 0.5 NT.

Attention Switching

The attention switching factor consisted of two levels representing alternate switching strategies by which the human operator can service vehicles with concurrent requests. Under the *First-in, First-Out (FIFO)* strategy, operators service vehicles on a first-come basis. In contrast, the *Highest Attribute First (HAF)* strategy relies on events having a specific criterion and the operator selects the vehicle with the highest priority value. In this experiment, priorities for vehicle servicing were as follows: The USV performing an ISR task was assigned the highest priority, followed by the UAV performing the ISR task, and the communications UAV had the lowest priority. Other priority-assignment schemes were not investigated in this experiment.

Simulation

Two simulation-based studies were performed for this effort. The first compared mission performance and operator utilization across the four vehicle teams for each of the three previously described Levels of Neglect (*Macro, Macro/Micro,* and *Micro*) for the *FIFO* Vehicle Switching strategy (a 4x3 study design). The second study compared mission performance and operator utilization for *heterogeneous* vehicle teams, for all three Levels of Neglect and for both the *FIFO* and *HAF* Vehicle Switching strategy (a 3x3x2 study design).

A total of 21 experimental treatments were completed to obtain data for this evaluation. Data was collected only once for identical conditions across the two studies, and was not collected for combinations of the HAF queuing scheme and the homogeneous team factor levels, as there is no difference between the *FIFO* and *HAF* strategies for a homogenous team.

Thirty simulation replications were conducted for each treatment condition, for a total of 900 runs. In each replication, data was collected for two dependent variables. The first dependent variable, *Mission Performance*, was a metric describing average weighted vehicle performance across the entire mission. The second dependent variable, *Operator Utilization*, was the ratio of total time the operator was busy servicing vehicles to total mission (simulation) time.

Models for this experiment were developed using the Arena® discrete event simulation modeling language, and simulations were run on a Fujitsu T4000 series tablet PC with a 1.80 GHz Intel Pentium processor.

RESULTS AND DISCUSSION

Because the discrete event simulation used in this evaluation lacked the variance expected for human subject data, a family-wise significance level of 0.001 was used wherever appropriate.

Study 1: Team Heterogeneity and Level of Neglect

The first of these simulation studies compared mission performance and operator utilization across multiple levels of team heterogeneity and level of neglect, with the attention switching factor fixed at the *FIFO* level (the 4x3 study).

A 4x3 MANOVA (team-heterogeneity x LON) revealed through the Wilk's Lambda test significant main effects for both factors ($p < 0.0001$), as well as a significant two-factor interaction ($p < 0.0001$). It was also determined through a univariate analysis that the two-way interaction was significant for both mission performance and operator utilization ($p < 0.0001$). The next step was therefore to compare simple effects for each of these DVs.

The box plots for the performance DV are presented in Figure 1. For the treatment means that had the smallest differences between them, nine simple contrasts using the Bonferroni procedure were conducted in order to check for significance. Using the experiment-wise adjusted error of 0.0001 (0.001/9), all mean contrasts were significant ($p < 0.0001$), except for the contrasts between the *Team3* treatment means for *Macro* and *Macro/Micro* LON (p-value = 0.0003), as well as the *Macro* treatment means for *Team1* and *Team3* (p-value = 0.0006).

The box plots for the utilization dependent variable are presented in Figure 2. For the treatment-mean contrasts that were the smallest (Team1/Micro and Team4/Micro; Team2/Macro and Team4/Macro; Team2/Micro and Team4/Micro), three simple contrasts using the Bonferroni procedure were conducted in order to check for significance.

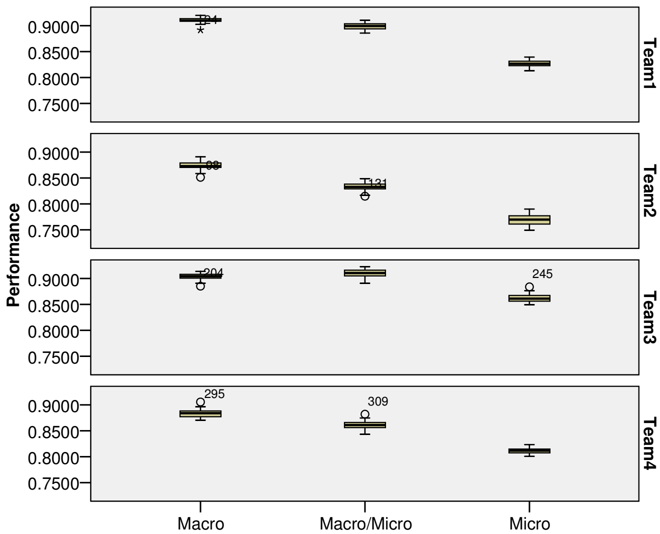

Figure 1. Performance by LON for each team

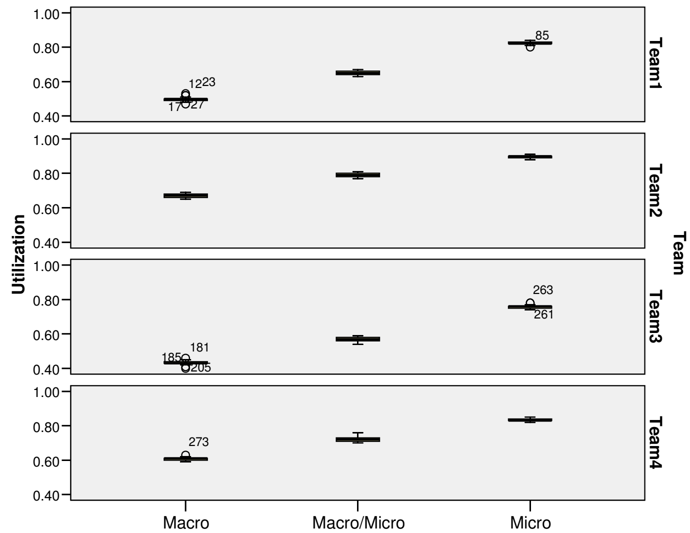

Figure 2. Operator Utilization by LON for each team

Using the experiment-wise adjusted error of 0.0003 (0.001/3), all contrasts resulted in significant differences ($p < 0.0001$), with marginal significance for the contrast between Team1/Micro and Team4/Micro ($p = 0.0003$).

The first experiment showed that the effect of heterogeneity in UV teams on system performance and operator utilization depends on the type of heterogeneity present. When heterogeneity was created by replacing one of the vehicles in a homogeneous team with another vehicle/task pair that had a smaller NT (*Team2*), performance decreased and operator utilization increased. This can be attributed to the fact that the lower overall NT in *Team2* created an increase in operator utilization, which likely led to increased wait times and degraded performance. *Team2* also experienced context switching times, which did not exist for the homogeneous team, and this further exacerbated the performance reduction.

However, when heterogeneity was created by replacing one of the vehicles in the homogeneous team with another vehicle/task pair that had a larger NT (*Team3*), the results were different. When operators had a *Macro* level of neglect strategy, there was no statistically significant difference in performance between the two teams. Although, the introduction of the high NT vehicle/task pair significantly reduced operator utilization for *Team3*, the drop in utilization was not enough to counteract the context switching times experienced when supervising *Team3*. However, under more *Micro* LON strategies, the utilization drop was even larger when supervising the heterogeneous team, which created significantly better performance. This suggests that by introducing certain forms of heterogeneity that increase the average NT of the team, it is possible to reduce utilization and even increase performance at certain levels of neglect.

Finally, *Team4*, which had the greatest heterogeneity across NTs, yielded both significantly higher utilization and lower system performance. One explanation for this result is that although the average NT across vehicles was similar for *Team1* and *Team4* (60s and 70s, respectively), heterogeneity was only present in *Team4*, which resulted in lower performance due to the existence of context switching times across different vehicles. Deeper analysis will be necessary, however, to determine whether the size of the spread in NTs across vehicles has any effect on the increased utilization and reduced performance.

It is also important to note that for all four teams, when going from a *Macro* to a *Macro/Micro* strategy or from a *Macro/Micro* to a *Micro* strategy, average operator utilization increased significantly, as expected. It was also the case that the increase in utilization tended to be accompanied by a significant drop in system performance. There are two likely explanations for this behavior. First, increased operator utilization likely led to reduced situation awareness (or increased wait time due to loss of situational awareness). Second, the increased rate of interaction with vehicles in the more *Micro* interaction levels resulted in saturated operators. This, in effect, created a large increase in queuing wait times. Both increased queuing wait times and wait times caused by loss in situation awareness are detrimental to performance, causing vehicles to be serviced at periods greater than their assigned NT.

Study 2: Team Heterogeneity and Switching Strategy

The second study compared mission performance and operator utilization across multiple team-heterogeneity, level of neglect, and attentional switching strategy. A 3x3x2 MANOVA (team-heterogeneity x LON x attention switching) revealed through the Wilk's Lambda test significant main effects for all three factors ($p < 0.0001$). The test did not reveal significance for the three way interaction ($p = 0.4176$), the LON x switching 2-way interaction ($p = 0.0334$), or the team- heterogeneity x switching 2-way interaction ($p = 0.0107$). There was, however, a significant team-heterogeneity x LON 2-way interaction ($p < 0.0001$). A univariate analysis revealed that the team-heterogeneity x LON two-way interaction was significant for both DVs ($p < 0.0001$), and that the switching main effect was significant only for the utilization DV (utilization: $p < 0.0001$; performance: $p = 0.0031$).

For the attention switching factor, it was found that under the *HAF* strategy, average operator utilization was significantly higher than under the *FIFO* strategy. This can be attributed to the fact that by moving high priority (low NT) vehicles to the top of the queue, average queuing wait times for the low priority vehicles increased (since they would always have to wait behind the high priority vehicles). Since, with the current model the length of required interaction time is proportional to the size of a vehicle's wait time, this resulted in longer interaction times and hence higher average operator utilization.

Although there was a significant increase in utilization, the results showed no significant change in performance under the *HAF* and *FIFO* switching strategies. Although it was hypothesized that the *HAF* strategy would increase performance by reducing the queuing wait times for the highest priority vehicles, this was not supported by the data. A possible explanation for this is that because the size of the vehicle teams in this experiment was small (only three vehicles), the reduction in queuing wait times was not likely to be substantial. For example, in the *FIFO*, strategy, the highest priority vehicle would, in the worst case, be waiting in a queue with 2 other vehicles ahead of it. If team size were larger, this reduction in wait time may be much more pronounced.

The important message from this analysis is that when considering the effects of alternate switching strategies, the number of vehicles in a team is an important consideration. When the team size is too small, it is unlikely that an *HAF* strategy will likely result in significant performance gains and could instead lead to a significant increase in average operator utilization.

CONCLUSIONS AND FUTURE WORK

This paper presents preliminary findings from a modeling and simulation effort exploring the impact of heterogeneity on the supervisory control of unmanned vehicle teams. Results demonstrated that when comparing heterogeneous teams to their homogeneous counterparts, the average potential neglect time across the vehicle team is decisive in predicting significant changes in mission performance and operator utilization. Heterogeneous teams with vehicle neglect times that are lower on average than those of a homogeneous team were likely to result in a significant increase in operator utilization.

On the other hand, heterogeneous teams with larger average neglect times across vehicles could cause a reduction in utilization and an increase in performance under certain operator interaction strategies. It was also noted that further investigation needs to address the effect of the size of the spread of neglect times across vehicles on average operator utilization and performance. Finally, the effect of varying operators' attention switching strategies was shown to be absent in the case of the small-sized vehicle teams examined in this study. Future work will involve using the lessons learned through this study to improve our model of operator interaction with unmanned vehicles teams and to make performance predictions that will be compared to findings from actual human-subject experiments.

ACKNOWLEDGEMENTS

The research was supported by the Office of Naval Research (ONR).

REFERENCES

Crandall, J. W., Goodrich, M. A., Olsen, D. R., & Nielsen, C. W. (2005). Validating Human-Robot Interaction Schemes in Multitasking Environments. *Systems, Man, and Cybernetics, 35*(4), 438-449.

Cummings, M. L., Nehme, C. E., Crandall, J., & Mitchell, P. J. (2007). Predicting Operator Capacity for Supervisory Control of Multiple UAVs. In L. Jain (Ed.), *Innovations in Intelligent UAVs: Theory and Applications*.

Goodrich, M. A., Quigley, M., & Cosenzo, K. (2005). *Task Switching and Multi-Robot Teams*. Paper presented at the 3rd International Workshop on Multi-Robot Systems.

Nehme, C. E., Crandall, J., & Cummings, M. L. (2008). *Using Discrete-Event Simulation to Model Situational Awareness of Unmanned-Vehicle Operators*. Paper presented at the 2008 Capstone Conference.

Nehme, C. E., & Cummings, M. L. (2007). *An Analysis of Heterogeneity in Futuristic Unmanned Vehicle Systems*. Cambridge, MA: Humans and Automation Laboratory, Massachusetts Institute of Technology.

Nehme, C. E., Mekdeci, B., Crandall, J., & Cummings, M. L. (in submission). The Impact of Heterogeneity on Operator Performance in Futuristic Unmanned Vehicle Systems. *The International C2 Journal, 1*(3).

Olsen, D. R., & Goodrich, M. A. (2003). *Metrics for Evaluating Human-Robot Interactions*. Paper presented at the Performance Metrics for Intelligent Systems.

Olsen, D. R., & Wood, S. B. (2003). *Fan-out: Measuring Human Control of Multiple Robots*. Paper presented at the CHI2004.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. *Presence, 11*(4), 335-351.

Sheridan, T. B. (1992). *Telerobotics, Automation, and Human Supervisory Control* Cambridge, MA: MIT Press.