Modeling the Impact of Operator Trust on Performance in Multiple Robot Control

Fei Gao, Andrew S. Clare, Jamie C. Macbeth, M. L. Cummings

Massachusetts Institute of Technology
Cambridge, MA 02139, USA
feigao@mit.edu

Abstract

We developed a system dynamics model to simulate the impact of operator trust on performance in multiple robot control. Analysis of a simulated urban search and rescue experiment showed that operators decided to manually control the robots when they lost trust in the autonomous planner that was directing the robots. Operators who rarely used manual control performed the worst. However, the operators who most frequently used manual control reported higher workload and did not perform any better than operators with moderate manual control usage. Based on these findings, we implemented a model where trust and performance form a feedback loop, in which operators perceive the performance of the system, calibrate their trust, and adjust their control of the robots. A second feedback loop incorporates the impact of trust on cognitive workload and system performance. The model was able to replicate the quantitative performance of three groups of operators within 2.3%. This model could help us gain a greater understanding of how operators build and lose trust in automation and the impact of those changes in trust on performance and workload, which is crucial to the development of future systems involving human-automation collaboration.

Introduction

Automated technologies are being implemented more and more in a variety of fields. While automation has the potential to reduce workload, increase efficiency and safety, it cannot totally replace humans in the system. Humans' advantages in terms of perception, flexibility, and adaptability make them essential to complex system operation. Increasingly, researchers are showing that pairing humans with automation works better than fully relying on automation (Anderson et al. 2000; Cummings et al. 2012; Ponda et al. 2011; Ryan 2011).

Researchers have been exploring the use of autonomous robots for Urban Search and Rescue (USAR) for over a decade. Robots can go to places that are impossible or too dangerous for human rescuers. In urban search and rescue, robots usually need to navigate through a complex environment to map the environment and look for victims. Currently in practice, two operators are usually required to manually control a rescue robot. With autonomous path planning, it is possible to reduce the workload of operators and increase the number of robots they can control.

Human trust in such autonomous path planners will be a crucial driver of performance when controlling multiple robots. Human trust in automation can be defined as the "attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee and See 2004). Operator trust in the automation can fluctuate due both to the operator's initial trust level and the behavior of the automation throughout the mission.

Both overtrust and undertrust in automation can be detrimental to system performance. Low human trust can be caused by automation "brittleness," in that the automation can only take into account those quantifiable variables, parameters, objectives, and constraints identified in the design stages that were deemed to be critical (Smith, McCoy, and Layton 1997). Also, "overtrust" in automation has been cited in a number of costly and deadly accidents in a variety of domains (Parasuraman and Riley 1997). Overtrust in the automation can lead to the phenomenon of automation bias (Mosier et al. 1998), where operators disregard or do not search for contradictory information in light of an automation-generated solution which is accepted as correct (Cummings 2004). Empirical studies have shown that when working with imperfect automation, automation bias can occur (Chen and Terrence 2009; Lee and Moray 1994).

Thus, this paper investigates the impact of operator trust in automation on performance when controlling multiple USAR robots. We begin by analyzing qualitative and quantitative data from an USAR experiment simulating a search and rescue scenario. Based on the findings from this analysis, we detail the development of a real-time human-automation collaboration, which can be used to model changes in operator trust over time and the impact on operator behavior and overall system performance. Finally, we exercise the model on the experimental data set to investigate its usefulness.

USAR Experimental Data Analysis

Our experimental data comes from a previous USAR experiment (Gao, Cummings, and Bertuccelli 2012) that was conducted using a 3-D simulation testbed based on USARSim (Lewis, Wang, and Hughes 2007). The task of the operators was to monitor the cameras of robots and mark the positions of victims on the map when they appeared in the cameras. The goal was to mark as many victims as possible correctly. By default, robots searched autonomously for victims based on a path-planning algorithm. Operators could choose to take manual control and teleoperate an individual robot during this process when they felt it was necessary. Operators worked in teams of two to monitor a total of 24 robots. Each team went through three trials of 25 minutes. In these three trials, the building maps were the same, but the locations of the 34 victims were different. At the end of each trial, subjective workload ratings were obtained from each operator using the NASA-TLX scale (Hart 1988), which measures six sub-dimensions of workload.

We conducted a number of analyses to investigate the impact of operator trust in automation on performance. First, as shown Figure 1, operators spent a longer time on teleoperation in later trials than in earlier ones within a single experiment. ANOVA analysis shows that this impact of trial sequence on the time spent on teleoperation is significant ($F(2,141)=7.37$, $p<0.001$).

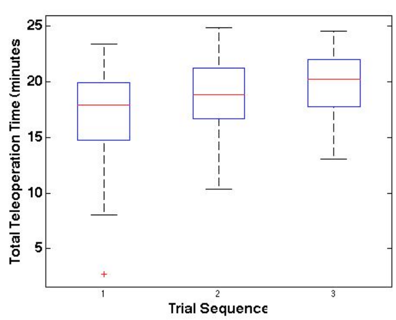

Figure 1: Total Teleoperation Time versus Trial Sequence

In interviews after the experiment, many participants said that the path-planning algorithm did not do a very good job and could not be trusted. They stated that robots often went back to places already explored, sometimes multiple times, while leaving some other places unexplored. They complained that the search was not thorough if they relied on the path planner only. Even though intervening via teleoperation requires more effort than just relying on the path-planning algorithm, operators chose to teleoperate when the robots were not going to places the operators wanted them to go.

Thus, both quantitative and qualitative data indicate that operators lost trust in the automation throughout the three trials and chose to intervene more frequently. This indicates a link between the operator's perception of the automation's capability, operator trust, and the amount of teleoperation conducted by the operator. The data also show that trust can change both during and in-between missions.

In a second analysis, for each 25-minute experiment trial, we calculated the number of teleoperation "actions" per minute. To smooth out short-term fluctuations in the time series data, we took the moving average over each five-minute period, resulting in 21 data points per trial. We then utilized hierarchical clustering to classify each of the 144 experiment trials into groups. Our goal was to identify groups of operators who had distinct behavior in terms of the frequency of teleoperation.

This analysis identified six distinct clusters of trials. The first three clusters contained only seven trials in total, and were removed from further analysis. The last three clusters had different levels of teleoperation as shown in Figure 2a. The trials in Cluster 4 (named Low TeleOp), shown in red squares in Figure 2a, had the lowest frequency of teleoperation. These trials also had significantly worse performance (($F(2,134)=16.67$, $p<0.001$), in terms of total victims found, than the other two clusters, as shown in Figure 2b. Thus there is a positive relationship between decreased teleoperation frequency in this experiment and decreased performance. Combined with the previously discussed link between trust and the frequency of teleoperation, it indicates an indirect, but crucial, relationship between operator trust in automation and system performance.

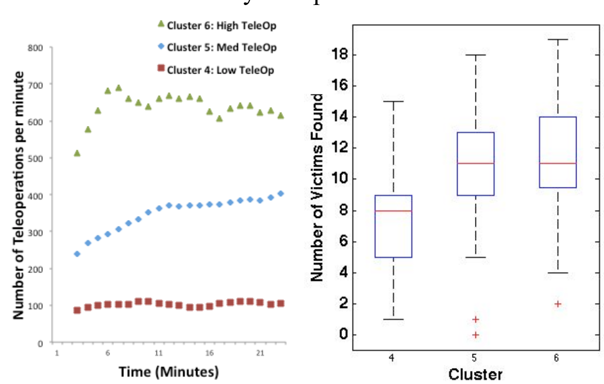

Figure 2: (a) Average Teleoperation Frequency versus Mission Time; (b) Number of Victims Found by Teleoperation Cluster

It should be noted, however, that while the trials in Cluster 6 (named High TeleOp) had significantly more teleoperations than those in Cluster 5 (named Med TeleOp) (Figure 2a), there were no statistically significant differences in system performance between these two groups

(Figure 2b). There appear to be diminishing returns in terms of performance with more teleoperation.

Finally, in order to investigate whether the frequency of teleoperation is related to operator workload, we analyzed the differences in NASA-TLX workload ratings among the three clusters. ANOVA analysis showed that there were significant differences on the temporal demand dimension of workload between clusters ($F(2,134)=68.37$, $p<0.001$), but not on overall workload ratings or any other dimensions. Operators in the clusters with higher amounts of teleoperation reported higher temporal demand, which indicates that they felt increased time pressure and that the pace of the simulation was rapid. This indicates that operator workload is related to the number of operator interventions.

Human-Automation Collaboration Model

Based on the findings from the above data analysis, we developed a computational model that can simulate the control of multiple robots throughout a hypothetical mission. In contrast to purely theoretical or conceptual models, such a computational model leverages computer simulations to both promote deeper understanding of human operator performance and provide testable predictions about human behavior under different circumstances (Gao and Lee 2006).

A number of computational models have been developed previously to simulate human-automation collaboration for controlling multiple robots (Cummings and Mitchell 2008; Nehme 2009; Olsen and Wood 2004; Rodas, Veronda, and Szatkowski 2011; Savla et al. 2008). The model presented here builds on this previous work in two ways. First, in contrast to these previous models, this model explicitly simulates the human-automation collaboration required for *goal-based* control (Clare and Cummings 2011), where a team of semi-autonomous robots conducts a mission under the command of an automated planning algorithm until the human operator intervenes. Second, this model is the first to leverage system dynamics modeling techniques (Sterman 2000) for simulating a human-automation collaborative system for controlling multiple robots. System dynamics methods enable the model to utilize both qualitative and quantitative variables, capture non-linear human behavior and performance patterns, and model the impact of latencies and feedback relationships among trust, workload, and performance.

The model simulates a human operator, a team of robots, and the associated automation at an abstract level, yet provides concrete metrics such as system performance, the frequency of operator interventions, and operator workload throughout the mission. Although the seed data comes from a team-based experiment, this model focuses on the concept of a single operator controlling multiple robots, thus team coordination and task allocation among multiple operators is not considered. We leave this for future work.

The model implements a set of equations which are calculated at discrete time steps using the Vensim® simulation software package. In the next section, we describe the model in detail.

Model Description

The model, as shown in Figure 3, consists of three major components: the System Performance module, the Trust in Automation feedback loop, and the Cognitive Overload feedback loop.

First, in order to properly model real-time human-automation collaborative control of a team of robots, an effective, yet simple model of system performance is necessary. The System Performance module is inspired by the diffusion model, which has been used to model the spread of new ideas, the adoption of new products, or the spread of contagious diseases (Sterman 2000). In the USARSim experiment, once a victim is visited, it is marked as "found", which is one performance metric. As more and more victims are found, the likelihood that a new victim is found declines, which decreases the Victim Discovery Rate.

There is an initial base search speed for the team of robots, which is a system parameter dependent on the number of robots, speed of the robots, and camera ranges of the robots. Finally, the human contribution to the human-automation collaboration is represented in the model as an additive Human Valued Added factor to the Victim Discovery Rate, as shown in Figure 3.

The second major component of the model is the Trust in Automation feedback loop, as shown in Figure 3. The trust loop of the model draws from a previous computational model of human trust in automation (Gao and Lee 2006) and has been adapted for modeling human-automation collaboration for control of multiple robots. We first assume that the operator perceives the capability of the automation as a ratio of the current rate of discovering new victims over the operator's expected rate of discovery of victims. Thus, if the system is behaving exactly as expected, the capability of the automation would be 100%.

The model takes as an input parameter the operator's initial trust level, which can vary widely based on the operator's prior knowledge, past experiences, and training (Lee and Moray 1994; Moray, Inagaki, and Itoh 2000). Trust is often dynamic, however, and can fluctuate throughout a mission based on the operator's perception of how well the automation is performing (Lee and Moray 1992; Muir and Moray 1996). A number of studies have found that human trust has inertia, where automation errors do not necessarily cause instantaneous loss in trust, but recovery in trust from severe failures can also be slow (Lee and Moray 1994; Lewandowsky, Mundy, and Tan 2000; Parasuraman 1993). In addition to the literature on trust,

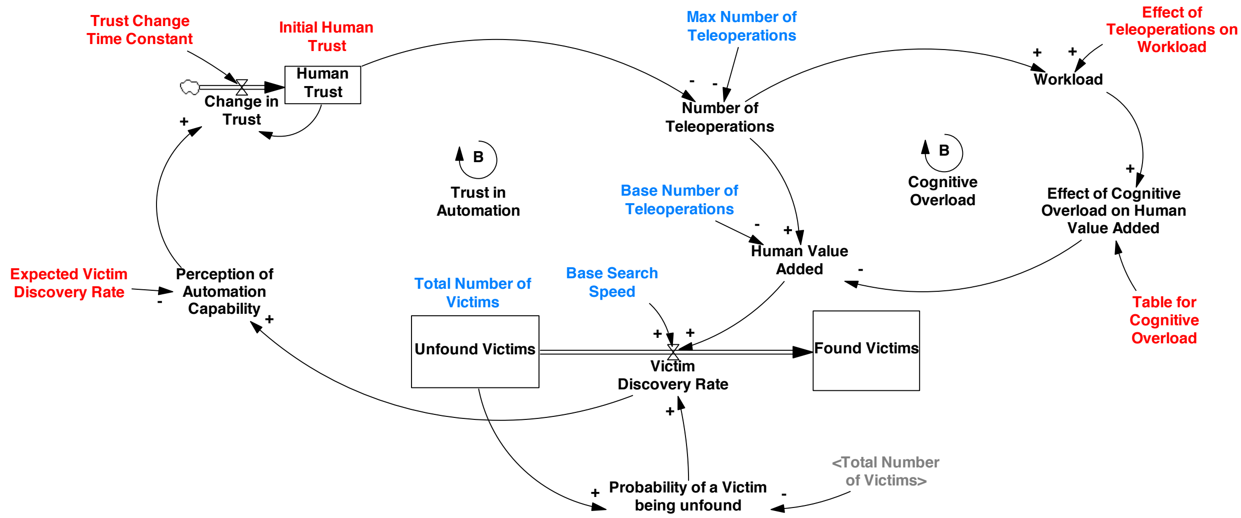

Figure 3: Human-Automation Collaboration Model

our experimental data analysis showed what appeared to be a link between perception of automation capability, human trust, and the frequency of teleoperation. To reflect the dynamic nature of trust, the model adjusts the operator's trust to the operator's perception of automation capability with a time delay determined by a Trust Time Change Constant that is representative of the operator's trust inertia.

Based on the findings from the experimental data analysis, we implement two additional features in the model to complete the Trust in Automation feedback loop. First, we model the frequency of teleoperation of the robots as inversely dependent on Human Trust level (higher trust means less likely to intervene). Second, a higher number of teleoperations improves the value the human operator adds to the team of robots, which increases the rate of finding victims.

The third major component of the model is the Cognitive Overload feedback loop. It has been shown in numerous previous studies that human cognitive workload has a significant impact on both human and system performance (Clare and Cummings 2011; Cummings, Clare, and Hart 2010; Cummings and Nehme 2010). As theorized in the Yerkes-Dodson curve (Yerkes and Dodson 1908), up to a certain point, increased workload can be beneficial to performance. Once the operator reaches what we refer to as cognitive overload, performance begins to suffer.

Thus, the model performs a simple calculation of cognitive workload by scaling the number of teleoperations into a workload value between 0% and 100%. Then the Effect of Cognitive Overload on Human Value Added to the search process is calculated based on a table function. This is consistent with the "Burnout" formulation used in previous system dynamics models (Sterman 2000). It has been established in previous literature that a utilization level over 70% can lead to performance decrements (Cummings and Guerlain 2007; Nehme 2009; Rouse 1983; Schmidt 1978). Thus, up to a workload level of 70%, there is little change in the effect of workload. Above a workload level of 70%, however, the effect would cause a drop in Human Value Added.

Finally, we close both feedback loops in the model by relating the number of teleoperations and the operator's workload back to system performance, as shown in Figure 3. The Human Value Added to the search process is calculated by the number of teleoperations divided by a base number of teleoperations, multiplied by the effect of cognitive overload. If the operator performs more teleoperations than the base amount, there will be a positive effect on victim discovery performance, which is consistent with our experimental data analysis. However, if the operator's cognitive workload reaches too high of a level, there will be a decrease in Human Value Added. This represents the fact that the rate of operator intervention and the effectiveness of these interventions can be in tension, as previous research has shown that high rates of human intervention in a highly automated multi-robot system can lead to worse performance (Clare, Maere, and Cummings 2012; Cummings, Clare, and Hart 2010).

As with any model, we made a number of assumptions in the development of the model. In the next section, we aim to test these model assumptions by using the model to simulate the behavior of different operators.

Experimental Data for Model Validation

We used the model to simulate the average behavior of the operators in each of the three clusters representing low, medium, and high frequency of teleoperation. After generally fitting the model to the average behavior of all operators, we were able to simulate the differences in behavior and performance of the three clusters by modulating only two parameters: the initial trust level and the trust change time constant. Based on the qualitative feedback from operators, we made the assumption that the lowest frequency of teleoperation group, Low TeleOp, had both a high initial trust level and the highest trust inertia (longest trust change time constant), as they were least likely to adjust their trust level to evidence that the automation was suboptimal. The opposite was true for the highest frequency of teleoperation group, High TeleOp, which was modeled with a low initial trust level and low trust inertia, as they were willing to adjust their trust level to their perception of automation capability. The parameter changes that were selected for each cluster are shown in Table 1. These parameter values were determined empirically by seeking a good fit to the data for each group.

The simulation output for Number of Teleoperations is compared to average experimental data ±1 Standard Error (SE) for each cluster in Figure 4, with the R^2, Root Mean Square Error (RMSE), and p-values of the Pearson correlation coefficient shown in Table 2.

Cluster	Initial Trust Level	Trust Change Time Constant
High TeleOp	15%	1800s
Med TeleOp	60%	3600s
Low TeleOp	95%	5400s

Table 1: Parameter Settings for Each Cluster

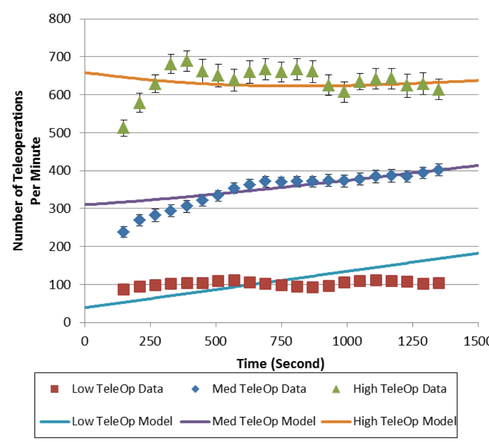

Figure 4: Number of Teleoperations: Simulations vs. Data ±1 SE

Cluster	R^2	RMSE	P-value
High TeleOp	0.412	42.79	0.002
Med TeleOp	0.809	25.62	<0.001
Low TeleOp	0.200	33.51	0.042

Table 2: Number of Teleoperations: Simulation to Experimental Data Fit

The best fit was achieved with the data from cluster Med TeleOp, as the model captured the steady increasing trend in the frequency of teleoperation. In the cluster High TeleOp, the fit was acceptable, but not ideal because of what may have been a transient at the start of the experiment where operators did not initially perform as many teleoperations as they did only minutes later when they achieved a steady state frequency of teleoperation. Removing the first two minutes from the High TeleOp data lowered the RMSE to 29.49, a 31% improvement in fit. Finally, while it may appear that cluster Low TeleOp had a subpar fit, with a coefficient of determination (R^2) of only 0.2, recall that R^2 is a measure of the proportion of the variation in the experimental data explained by the model. As there was little variation over time in the frequency of teleoperation for this group in the experimental data, we argue that the model cannot and should not recreate these small variations in order to avoid overfitting.

The simulation output for Found Victims is compared to average experiment data ±1 SE for each cluster in Figure 5. The simulations had a good fit to the experimental data, with the R^2, RMSE, and final value percent errors shown in Table 3. The model was able to calculate the average final number of victims found in each cluster within 2.3%. The performance curve for cluster Low TeleOp underestimates the number of victims found for much of the earlier portion of the mission, which may indicate that our System Performance module requires refinement, especially when modeling operators with low rates of intervention.

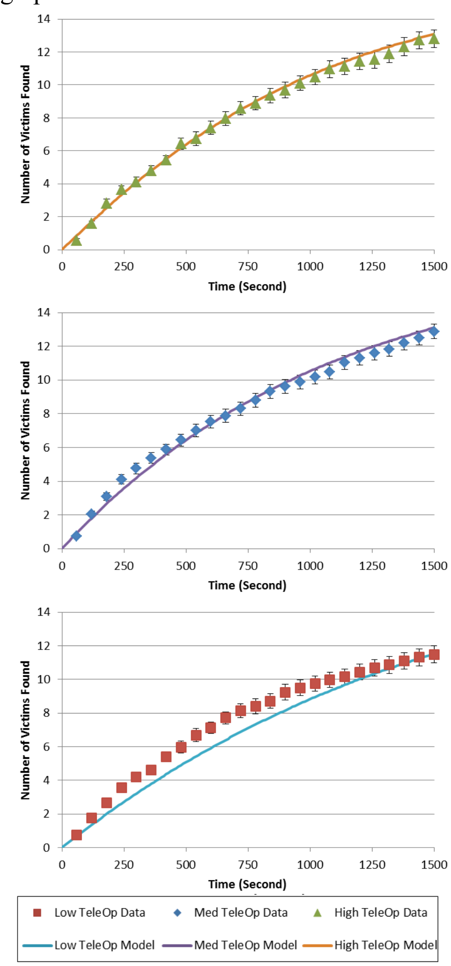

Figure 5: Found Victims: Simulations vs. Data ±1 SE

Cluster	R^2	RMSE	Percent Error at Finish
High TeleOp	0.998	0.228	2.3%
Med TeleOp	0.996	0.353	1.7%
Low TeleOp	0.985	0.839	0.4%

Table 3: Found Victims: Simulation to Experimental Data Fit

Overall, the model was able to replicate the behavior and performance of three groups of operators only by adjusting the initial trust level and the amount of trust inertia of each group. While we make no claims that this is the exclusive or even optimal set of variables that could be used to represent the differences among these groups of operators, the qualitative and quantitative experimental data indicated that these were the most likely differences among these clusters. Two variables provided enough predictive power while still maintaining a parsimonious representation of the world.

The lowest trusting group, High TeleOp, had a higher frequency of teleoperation, yet achieved roughly the same performance as the moderate trust group, Med TeleOp. The model replicated this behavior because the Cognitive Overload feedback loop counteracted the impact of additional teleoperation on the human value added to system performance.

There are many limitations to this simple model of human-automation collaboration. The model assumes a very simple model of the search process, which is not perfectly accurate as discussed earlier. The model assumes that operators can actually perceive the capability of the automation by observing the robots on a computer display. Also, the model simply assumes that the amount of teleoperation is the key driver of human value added, without considering any strategies for which robot to take control of at what time. Other assumptions include that there is no time delay in the perception of automation capability, operator expectations of performance are static, the number of interventions can be modeled as a continuous variable as opposed to discrete interventions, and human cognitive workload can be calculated by scaling the number of teleoperations directly to workload, all of which merit further investigation.

In addition to these simplifying assumptions, any simulation model such as this must be tuned to the specific situation that it is attempting to represent. The sensitivity of the tuning parameters for this specific USAR experiment has not yet been explored. Finally, the fit of the simulation to experimental data is acceptable, but not perfect, while additional refinement could push the model towards the danger of overfitting and a loss of generalizability. The fit of a model to a single data set is not sufficient for model validation. It does, however, demonstrate the potential usefulness of this model for describing and potentially predicting the behavior and performance of a human-automation collaborative system for controlling a team of robots.

Conclusion

We developed a system dynamics model to simulate the impact of operator trust on human and system performance in multiple robot control. Data analysis from the search and rescue experiment showed that operators spent more time on teleoperation in later trials than in earlier ones within a single experiment, demonstrating that they likely were losing trust as the missions went on. Operators who rarely conducted manual teleoperations had lower performance than operators who had moderate or high rates of teleoperation. Also, operators who more frequently conducted teleoperation reported higher temporal demand, which indicates that they felt increased time pressure and that the pace of the simulation was rapid. These findings led us to build into the model the assumption that more frequent human intervention improved system performance up to the point where the operator reached cognitive overload, at which point additional intervention did not improve performance.

In this model, trust and performance form a feedback loop, in which operators perceive the performance of the system, calibrate their trust, and adjust their control of the robots. A second feedback loop incorporates the impact of trust on cognitive workload and system performance. This model is consistent with what we observed in an experiment simulating a search and rescue scenario. We found that operators decided to manually control the robots when they lost trust in the autonomous planner that was aiding the operator.

Finally, the model was able to replicate the experiment data quantitatively. The model had mixed results in simulating the behavior of the three groups in terms of frequency of teleoperation, but was able to accurately replicate the final system performance of each group to within 2.3%. Overall, the model could accurately simulate the system performance of all three groups because it incorporated the impact of both trust in the automation and cognitive workload on system performance.

This model could help us gain a greater understanding of how operators build and lose trust in automation and the impact of those changes in trust on performance, which is crucial to the development of future systems involving human-automation collaboration. Operators adjust their trust based on the perception of system capability over time. Systems should be designed so that operators can develop a clear understanding of the system capability as soon as possible. At the same time, manual control should be accessible and easy to use, so that operators can intervene to compensate for suboptimal automation.

Acknowledgments

This research is sponsored by the Office of Naval Research and the Air Force Office of Scientific Research. We would also like to acknowledge Matthew Redmond and Radhika Malik for their programming contributions, and Hosea Siu and Victoria Palay for their help with conducting the experiment. Thanks to Professor Mike Lewis and his students at University of Pittsburg for providing the testbed. Thanks

to Professor Nelson Repenning for his assistance with system dynamics modeling techniques.

References

Anderson, D., Anderson, E., Lesh, N., Marks, J., Mirtich, B., Ratajczak, D., and Ryall, K. 2000. Human-Guided Simple Search. Paper presented at the Seventeenth Conference on Artificial Intelligence (AAAI-00), Austin, TX.

Chen, J. Y. C., and Terrence, P. I. 2009. Effects of Imperfect Automation on Concurrent Performance of Military and Robotics Tasks in a Simulated Multi-tasking Environment. Ergonomics, 52(8): 907-920.

Clare, A. S., and Cummings, M. L. 2011. Task-Based Interfaces for Decentralized Multiple Unmanned Vehicle Control. Paper presented at the AUVSI Unmanned Systems North America, Washington D.C., August.

Clare, A. S., Maere, P. C. P., and Cummings, M. L. 2012. Assessing Operator Strategies for Real-time Replanning of Multiple Unmanned Vehicles. Intelligent Decision Technologies, 6(3): 221-231.

Cummings, M. L. 2004. Automation Bias in Intelligent Time Critical Decision Support Systems. Paper presented at the AIAA 3rd Intelligent Systems Conference, Chicago, IL.

Cummings, M. L., Clare, A. S., and Hart, C. S. 2010. The Role of Human-Automation Consensus in Multiple Unmanned Vehicle Scheduling. Human Factors, 52(1): 17-27.

Cummings, M. L., and Guerlain, S. 2007. Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles. Human Factors, 49(1): 1-15.

Cummings, M. L., How, J., Whitten, A., and Toupet, O. 2012. The Impact of Human-Automation Collaboration in Decentralized Multiple Unmanned Vehicle Control. Proceedings of the IEEE, 100(3): 660-671.

Cummings, M. L., and Mitchell, P. J. 2008. Predicting Controller Capacity in Supervisory Control of Multiple UAVs. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 38(2): 451-460.

Cummings, M. L., and 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.

Gao, F., Cummings, M. L., and Bertuccelli, L. F. 2012. Teamwork in Controlling Multiple Robots. Paper presented at the ACM/IEEE International Conference on Human-Robot Interaction (HRI '12), Boston, MA.

Gao, J., and Lee, J. D. 2006. Extending the Decision Field Theory to Model Operators' Reliance on Automation in Supervisory Control Situations. IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans, 36(5): 943-959.

Hart, S., & Staveland, L. 1988. Development of the NASA-TLX: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human mental workload (1.0 ed., pp. 139-183v). Amsterdam: North Holland.

Lee, J. D., and Moray, N. 1992. Trust, Control Strategies and Allocation of Functions in Human-Machine Systems. Ergonomics, 35(10): 1234-1270.

Lee, J. D., and Moray, N. 1994. Trust, Self-Confidence, and Operators' Adaptation to Automation. International Journal of Human-Computer Studies, 40(1): 153-184.

Lee, J. D., and See, K. A. 2004. Trust in Automation: Designing for Appropriate Reliance. Human Factors, 46(1): 50-80. doi: 10.1518/hfes.46.1.50_30392

Lewandowsky, S., Mundy, M., and Tan, G. P. 2000. The Dynamics of Trust: Comparing Humans to Automation. Journal of Experimental Psychology: Applied, 6(2): 104-123.

Lewis, M., Wang, J., and Hughes, S. 2007. USARsim: Simulation for the study of human-robot interaction. Journal of Cognitive Engineering and Decision Making, 1(1): 98-120.

Moray, N., Inagaki, T., and 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.

Mosier, K., Skitka, L. J., Heers, S., and Burdick, M. D. 1998. Automation Bias: Decision Making and Performance in High-tech Cockpits. International Journal of Aviation Psychology, 8(1): 47-63.

Muir, B. M., and Moray, N. 1996. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics, 39(3): 429-460.

Nehme, C. 2009. Modeling Human Supervisory Control in Heterogeneous Unmanned Vehicle Systems. Ph.D. diss., Dept. of Department of Aeronautics and Astronautics, Massachusetts Institute of Technology.

Olsen, D. R., and Wood, S. B. 2004. Fan-out: Measuring Human Control of Multiple Robots. Paper presented at the SIGCHI conference on Human factors in Computing Systems, Vienna, Austria.

Parasuraman, R. 1993. Effects of adaptive function allocation on human performance. In D. J. Garland & J. A. Wise (Eds.), Human factors and advanced aviation technologies (pp. 147-157). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Parasuraman, R., and Riley, V. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors, 39(2): 230-253.

Ponda, S., Ahmed, N., Luders, B., Sample, E., Hoossainy, T., Shah, D., Campbell, M., and How, J. 2011. Decentralized Information-Rich Planning and Hybrid Sensor Fusion for Uncertainty Reduction in Human-Robot Missions. Paper presented at the AIAA Guidance, Navigation, and Control Conference, Portland, OR, Aug. 8-11.

Rodas, M. O., Veronda, M., and Szatkowski, C. 2011. Developing a Decision Tool to Evaluate Unmanned System's Command and Control Technologies in Network Centric Operations Environments. Paper presented at the The Third International Conference on Advanced Cognitive Technologies and Applications, Rome, Italy.

Rouse, W. B. 1983. Systems Engineering Models of Human-Machine Interaction. New York: Elsevier North Holland.

Ryan, J. C. 2011. Assessing the Performance of Human-Automation Collaborative Planning Systems. Masters of Science diss., Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology.

Savla, K., Nehme, C., Temple, T., and Frazzoli, E. 2008. On Efficient Cooperative Strategies between UAVs and Humans in a Dynamic Environment. Paper presented at the AIAA Guidance, Navigation and Control Conference, Honolulu, HI.

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. doi: 10.1109/TSMC.1978.4310003

Smith, P., McCoy, E., and 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(3): 360-371.

Sterman, J. D. 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. New York: Irwin/McGraw-Hill.

Yerkes, R. M., and Dodson, J. D. 1908. The Relation of Strength of Stimulus to Rapidity of Habit-Formation. Journal of Comparative Neurology and Psychology, 18: 459-482.