Interaction of Automation and Time Pressure in a Route Replanning Task

Kip Johnson, Liling Ren, James Kuchar, and Charles Oman

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

Abstract

Two experimental studies were performed to determine the value of several types of automation for time-pressured replanning tasks. Subjects modified waypoints of two-dimensional routes on a computer screen in response to a sudden change in the displayed environment while having access to one of four levels of automation. In addition to a baseline case with no automation, subjects were assisted with automation that either reduced hazard exposure, ensured meeting time-to-target and fuel constraints, or combined hazard avoidance with meeting time and fuel constraints. Time pressure was imposed by requiring a replan of the route within four different levels of time. The two studies examined different ranges of time pressure, ranging from four levels between 20 and 55 seconds or four levels between 20 and 125 seconds. Results show that the presence of automation was most beneficial in the most time-pressured cases, and that the value of automation decreased as more time was available to the subject. Mid-level automation resulted in more routing errors than either full or no automation cases. Subjects were reticent to deviate from highly automated route suggestions even when significant improvements were still possible.

Introduction

Complex, uncertain, and time-critical environments push the limits of human sensory and cognitive ability, driving the need for automated decision-support systems. While there is a clear need for automation that reduces human cognitive workload, designing for an effective automated decision-aiding system is a difficult task (e.g., Parasuraman & Riley, 1997; Sexton, 1988; Sarter & Woods, 1995). Unstructured and uncertain aspects, with multiple competing interests and goals, characterize these complex environments. Full and complete automation may not be appropriate or feasible for complex environments because automation often does not have access to or cannot accurately model relationships between all relevant information (Parasuraman & Riley, 1997; Scerbo, 1996). Rather, automation working in parallel with a human for decision-making tasks would be more appropriate (Taylor & Reising, 1998; Schulte, et al., 1999; Layton, et al., 1994; Aust, 1996). Decision-support systems should take advantage of the human's ability to make value and risk judgments in the face of competing factors that may constrain a problem's analytical solution. Therefore, some form of cooperation between human and automation is generally required.

There are countless examples of beneficial automation spanning fields including transportation, process control, battlefield management, and medicine. Unfortunately, there are also many examples of automation that negatively affected performance due to unexpected interactions with the human operators. A number of aviation accidents and in-flight upset incidents, for example, demonstrate the need for continuing improvement of feedback on the current mode of automation in civil transport aircraft (Parasuraman & Riley, 1997). Similarly, conflicting responses from Boeing 757 pilots have been observed on whether cockpit automation reduced or increased total workload (Wiener, 1988). One study of navigation expert systems demonstrated that automated information processing caused a loss in situation awareness and the degradation of manual skills (Endsley & Kiris, 1995). These examples give a glimpse of the main problems associated with the introduction of automation: complacency, skill degradation, increases in cognitive workload, and loss of situation awareness.

There is no one solid solution for the design of automation, and much effort has been expended toward understanding what the different types of automation are and how they may be best applied (Parasuraman et al., 2000). The design of automation has many issues to resolve: to automate or not, how much human-in-the-loop interaction should be required or permitted, and how intelligent or adaptive to design the automated system (Sexton, 1988; Scerbo, 1996). One area of interest is that of using automation in very time-constrained tasks. Research in this area has been quite limited to date and has focused, for example, on human monitoring of fault-management automation in thermal-hydraulic systems (Moray et al., 2000).

Automation and Decision-Making Timescales

In problems with very little time to make a decision (i.e., on the order of one second), the human's sensory, cognitive, and motor abilities are such that automation must take action directly if it is to have any impact at all – there is simply insufficient time for a human to sense, interpret, and act. Automation is valuable in these situations due to its ability to obtain and process information more quickly than the human. In slightly less time-constrained situations (on the order of several seconds), the human does have time to sense, make judgments, and act, but can still be assisted by automation that relieves some of the cognitive workload. This is the principle behind alerting systems that provide commands to the operator. In less time-constrained situations, decision-aiding automation may be correspondingly less beneficial because the human's cognitive abilities no longer limit the quality of the solution. Additionally, a human can obtain and assess more information as time pressure decreases, generally opening up problems beyond easily automated tasks to those dealing with value judgments among different competing constraints. These are aspects best suited to human decision-making; thus, one might expect a spectrum of types of automation to be most appropriate as the time pressure on decisions is changed.

This interaction between automation and time pressure formed the foundation for two experimental studies. The primary experimental goal was to determine the relationships between varying degrees and types of automation assistance and time pressures as inputs, and the resulting human decision performance for an in-flight replanning task as outputs. The results from these studies will be used to help generate a generalized model for decision-support systems used in complex and time-critical environments.

Experimental Method

The experiment used a military combat flight environment theme, focusing on an in-flight replanning task. Replanning is a complex, time-critical, and highly dynamic task, rich with information input and output sources. The simulated missions were restricted to constant altitude and constant velocity flight of conventional military aircraft, with primary functions including air-to-ground, close air support, forward air control, and multi-role. Such aircraft could include, for example, the F-16 C/D Fighting Falcon, F-18 Hornet, A-10 Thunderbolt II, and the AC-130 H/U Gunship.

In the experiment, subjects observed a desktop computer screen that depicted a planform view of an aerial map containing the flight route and threats such as missile sites (see Fig. 1). The subjects interacted with the current route through waypoint manipulation following two primary replanning goals: 1) to first satisfy mission constraints, and then 2) to minimize the route cost. Mission constraints included avoiding the most severe threat level, arriving at a target point within an acceptable time-to-target (TTT) window, and having sufficient fuel at an egress (or exit) waypoint. Route cost was a function of the time of hazard exposure and deviation from the TTT goal, respectively modeled linearly and exponentially.

Threats were shown as irregular polygons in four different colors, representing different hazard levels. Threat severity increased in color from yellow, to orange, to red, and then to brown. The current and modifiable route was blue with star-shaped icons representing route waypoints. The route connected in serial order an entry (or start) point (upper right corner of Fig. 1), an intermediate rendezvous point (upper left in Fig. 1), the target (bottom right in Fig. 1), and an egress (or exit) point (bottom left in Fig. 1). Bar gauges (denoted #3 & #4 in Fig. 1) provided real-time TTT and fuel information. The fuel gauge showed the predicted fuel level at the egress point. As the route was lengthened, the fuel level at the egress point dropped linearly. The TTT gauge depicted the predicted arrival time at the target relative to a goal time and an acceptable window around that goal. As the length of the route before the target was lengthened, the TTT pointer would move to the right linearly. Lengthening a portion of the route after the target did not affect the time-to-target but did affect the fuel gauge. Accordingly, there was a challenging balance required between arriving at the target within the desired window (by lengthening or shortening the route between start and target) and not exceeding the fuel limits.

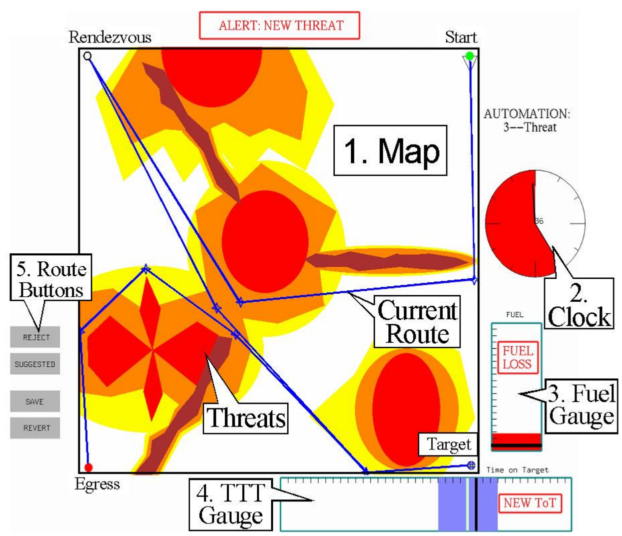

Figure 1. In-Flight Replanner ***Display***

After becoming familiar with the situation a sudden change in the environment was triggered and data collection began. Changes included pop-up threats, a shift in the time-to-target window, and a new fuel level requirement. When the situation update occurred, varying levels of automation in three test conditions provided a new route suggestion. In a fourth test condition (the control case) no new route suggestion was provided. Additionally, a countdown clock appeared on the screen indicating the time remaining to complete a final route plan (#2 in Fig. 1). While the clock counted down, the subject manually moved the route waypoints by clicking and dragging with a computer mouse. As the route was modified, the initial suggested route was continually displayed as a magenta dashed line for reference. The subject could also save or restore routes as he or she worked by clicking on dedicated screen buttons.

At the expiration of the time pressure shown on the clock, the scenario was ended and the computer scored the route. The route cost was determined based on the weighted distance through each of the four levels of threat, plus a component that depended on the actual time-to-target relative to the desired window. Fuel remaining at the egress point was also monitored and noted separately. At the end of each run, subjects were given feedback as to whether they satisfied threat, fuel, and time-to-target constraints.

Independent and Dependent Variables

Four route automation assistance categories were evaluated: None, Constraint, Hazard, and Full. The None condition did not provide any route suggestion beyond the original route that was displayed before the situation update occurred. Constraint automation provided a route suggestion that satisfied only the time-to-target and fuel constraints by extending the length of the route near the entry point as needed. It did not make any additional changes to the route to avoid hazard regions. Hazard automation shifted the original route to locally reduce the distance flown through threat regions. These routes might still violate the TTT or fuel constraints. Finally, in the Full automation condition both the Constraint and Hazard automation components were activated, producing a route that satisfied the time-to-target and fuel constraints and reduced the threat exposure. Even in the Full automation case, however, the route cost could be improved significantly from the starting plan generated by the automated route suggestion. This represented the fact that automation generally cannot observe all the required information and the human may still be able to make improvements.

The time pressure values included four conditions in each experiment. In the first experiment, the clock on the screen (denoted #2 in Fig. 1) counted down beginning at 20, 28, 40, or 55 seconds. In the second experiment, the clock counted down from 20, 55, 90, or 125 seconds, providing a cross-check of performance at low times and also opening up the study to slightly longer timescale decision-making. In four scenarios in each experiment, the subjects were allowed to continue to improve the route plan after the time allotment had expired and their route was scored. This allowed them to spend as much time as needed to generate their best plan. This optimal plan was then used in data analysis as a baseline for comparison by representing the best performance each subject could attain on a given map.

Sixteen scenarios were used in each experiment in a within-subject repeated measures experimental design. A four-by-four Graeco-Latin Square test matrix counterbalanced type of automation assistance, time pressure, and maps. Each subject ran through the 16 data collection scenarios in the same order. Each experiment used the same four base scenario maps of similar complexity, each one rotated four times by 90 degrees to generate the 16 effective scenarios. In post-experiment questioning, subjects indicated they did not recognize the similarity of the four rotated maps. The design of the maps was carefully controlled in an attempt to make them as similar as possible in terms of route complexity, yet different enough that familiarity with the best solution did not grow as the experiment proceeded.

The combination of map and automation level was such that all cases using None, Constraint, or Hazard started with some form of route failure (crossing the highest level threat, arriving at the target outside the time-to-target window, or arriving at the egress point without the required fuel level). Accordingly, one of the first tasks for the subjects in these cases was to correct these route failures. In the Full automation case, all scenarios began with a route that satisfied all the mission constraints.

Dependent variables included the route cost at the expiration of the time pressure, the number and type of constraint violations, the number of route modifications made by the subject, and subjective comments from questionnaires during and after the experiment.

Before beginning data collection, subjects were extensively trained using the computer interface to replan routes. They also received specific training on each of the automation levels and had opportunities to replan under each of the tested time pressure conditions. During initial training, the subjects could view an additional display that indicated the current route cost. The route cost display allowed them to gain a better understanding of the relative contribution to cost of flight through a hazard region vs. deviating from the TTT goal. The route cost display was not shown during data collection, however, since it represented an additional layer of automation that was beyond the scope of this study. To prepare the subjects for the lack of a route cost display, the subjects also completed several runs duringlater stages of training while using the same display configuration as the data collection runs.

Results

Fourteen graduate students participated in the first experiment, and eleven subjects (from a different pool of graduate students) participated in the second experiment. A repeated measures analysis was performed using a mixed regression and a two-way analysis of variance (ANOVA). The mixed regression best matched the Latin-Square design, and the ANOVA allowed for easy contrasts. Route costs were normalized for each subject against their minimal cost route (achieved from scenarios without time pressure) using the corresponding map. This enabled a direct comparison of each subject's rerouting performance under time pressure relative to their best possible routes under no time pressure with the same map.

Figure 2 provides a summary of the route costs for all subjects in both experiments. The vertical axis shows the mean log-normalized route cost across all subjects, while the horizontal axis shows the time pressure. Standard error of the mean is indicated by the error bars. The normalized route costs were generated by first dividing a subject's route cost for a map with time pressure by their best route cost for the same map when not under time pressure. The natural logarithm of this normalized cost was then taken to transform the cost data into an approximately normal distribution. Accordingly, a route cost of 0 in Fig. 2 corresponds to a route with the same quality as one developed with no time pressure. Increasing costs represent decreased replanning performance relative to this baseline value. Each increment of 0.1 in route cost in Fig. 2 is equivalent to approximately one minute of flight time through a second severity level hazard.

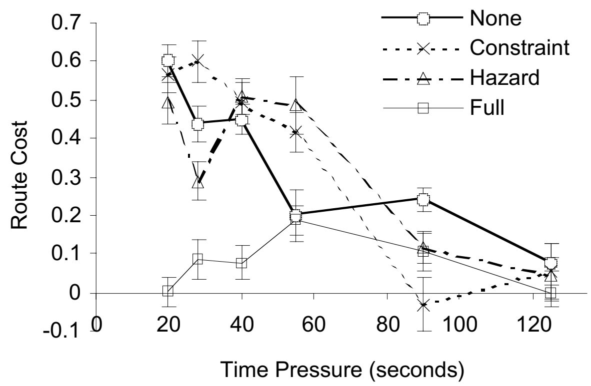

Figure 2. Time and Automation Interaction Results

Figure 2 first shows that performance in the None condition did significantly improve as more planning time was available, reducing the mean cost from 0.60 at 20 seconds to 0.07 at 125 seconds.

Performance using partial automation levels (Constraint and Hazard) was more complex relative to None. At 20 seconds, any automation level out-performed None, though the differences between Constraint and None or Hazard and None were not significant. At 55 seconds, None significantly out-performed Constraint or Hazard automation, $p<0.005$ with a paired t-test. At 125 seconds, there was no significant difference between None, Constraint, or Hazard automation.

Full automation likewise elicited an interesting and complex relationship with time pressure. At 20 seconds, subjects were able to perform as well with Full automation as they did when there was no time pressure on the same map. At 55 seconds, however, performance with Full (mean of 0.19) was significantly worse than at 20 seconds, for reasons that are explained below, $p<0.005$ with a paired t-test. In fact, at 55 seconds there was no significant difference between subjects using None and subjects using Full automation. At 125 seconds, there was no significant difference between Full and the other automation conditions tested. This trend suggests that automation loses value as time pressure is relaxed.

Figure 3 shows the rate of route failure for both experiments. A route was defined to have failed if it either (1) intersected a highest-level threat region, (2) did not arrive at the target within the acceptable time window (shown as #4 in Fig. 1), or (3) arrived at the egress point with less than the minimum amount of required fuel. The failure rate is defined as the fraction of all scenarios in which failures occurred in a given condition. As a reference, if subjects made no changes to the suggested route the failure rates would have been 100% for None, Constraint, or Hazard, but 0% for Full.

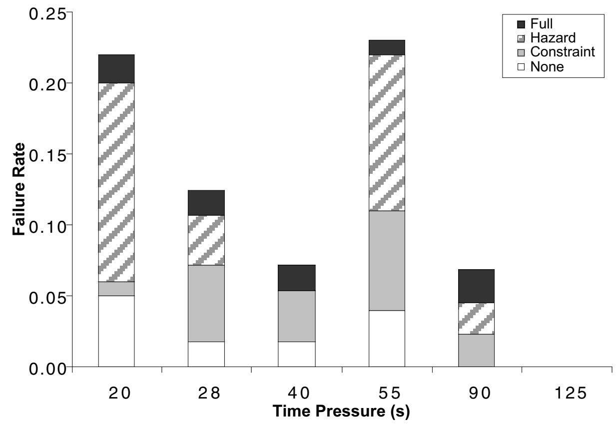

Figure 3. Total Failure Rate vs. Time Pressure

As Fig. 3 shows, no failures occurred at 125 seconds, while some degrees of failure occurred at other time pressure conditions. There was no significant difference between overall failure rates at 20 or 55 seconds. At 40 and 90 seconds the overall failure rates were significantly less than at 55 seconds. The rise in overall failure rate between 40 and 55 seconds (by a factor of three) appears to be due to a critical shift in subject behavior. When given 40 seconds or less, subjects generally made small, local improvements to the route in an attempt to resolve mission constraints. This is explained by the significant reduction in overall failure rate between 20 and 40 seconds with only minor changes in route cost (referencing Fig. 2). It appears that at 55 seconds, subjects felt they had enough time to attempt larger, more global route changes (for example rerouting in a different direction around a threat region). Often, however, the subjects apparently did not actually have enough time to successfully accomplish this global route change. When 90 or more seconds were available, such global route changes generally could be made, and both route cost and failure rate dropped.

It is also worth noting that the failures that occurred in the Full automation condition were entirely induced by the subjects. The scenarios with Full automation all began with a route that satisfied the mission constraints, so any failure that did occur was due to an error by the subjects.

Figure 4 shows a view of the overall failure rate as a function of the type of automation being used. Most of the failures occurred with either the Constraint or Hazard levels of automation, suggesting that incomplete automation may actually induce failures beyond a condition with no automation. If the failure rate is treated as a binomial process with some probability, then each of the differences in Fig. 4 is statistically significant, $p < 0.01$.

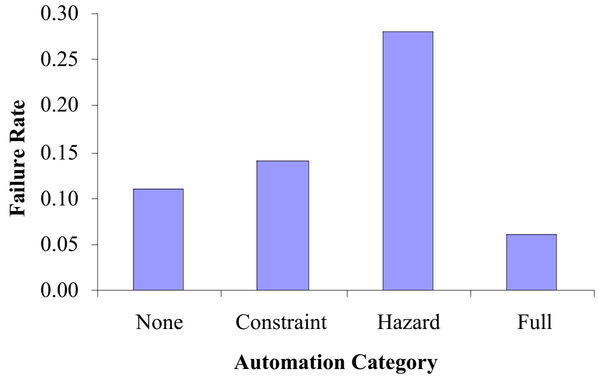

Figure 4. Failure Rate vs. Automation Level

Finally, Figure 5 shows the rate at which route changes were made by subjects during the second experiment. The mean rate of route modifications under Full automation (average of 0.19 per second) was significantly less than under the other conditions (average of 0.28 per second). This implies some degree of complacency with Full automation even though significant improvements in route cost could often be made. Route modification rates between None, Constraint, and Hazard were not significantly different.

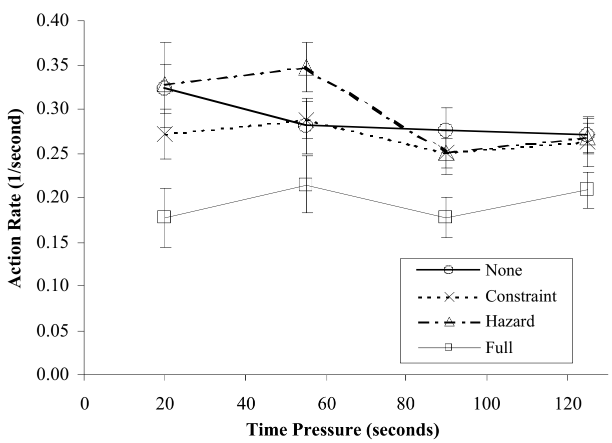

Figure 5. Route Modification Rates

Discussion

Based on the route cost metric, Full automation provided more benefit at higher time pressures; at less-constrained time pressures, performance with None was similar to performance with any type of automation. In addition, at short timescales Full automation provided more benefit than the sum of its individual Constraint and Hazard subcomponents. This implies there is a compounding benefit from automation that more fully integrates information. It is interesting to note, however, that subjects with Full made fewer route modifications than in other conditions with automation, even though significant route improvements could still be made. This suggests that good automation may induce some degree of complacency.

Automation can certainly provide a major benefit in time-pressured decision-making. However, the use of partially integrated information may induce human errors not otherwise observed without the decision-aiding automation. Trends suggested that Constraint and Hazard automation could actually be *detriments* to subject performance. There was a higher mission failure rate with Constraint (with failures in 14% of the scenarios) and Hazard (28%) than with None (11%). Full automation (6%), however, kept failure rates at levels lower than observed with None.

Performance trends suggest that replanning performance does not monotonically improve with increasing time available. From 20 to 55 seconds, performance with Full automation significantly decreased. From 55 to 125 seconds, performance with Full significantly improved. In addition, mission failures tripled from cases at 40 seconds (7% failure rate) to 55 seconds (23%). This deserves further investigation. At this point, however, it appears that the following observations may support the performance data. To achieve an acceptable plan, subjects needed to perform three cognitive subtasks. First, mission constraints (high-level threat, time-to-target, and fuel) needed to be satisfied or else the entire plan would be unacceptable. Second, small, local modifications to the route could be made to reduce threat exposure or improve the time to target. Third, major changes in the route were generally necessary to significantly improve route cost from the initial automated route suggestion. For example, the subject might significantly improve the overall route by flying an entirely different route around a threat region. These subtasks could be interleaved or performed in different orders or iteratively depending on the amount of time available.

In general, the three subtasks (constraint satisficing, local improvements, and global modifications) each required different amounts of time to complete. Achieving significant improvements in constraint satisfaction was possible within 20 seconds; this appeared to be one of the subjects' first subtasks because they made 80% of the initially unacceptable routes acceptable. From 20 to 40 seconds, subjects also generally made small improvements to routes to locally reduce route costs. At 55 seconds, however, many subjects began making large changes to the initial routes in an attempt to find the global minimal cost route before then making local adjustments. This subsequently required more time than anticipated to complete successfully, as evidenced by a significantly higher rate of mission failures at 55 seconds than at 40 seconds. Beyond 55 seconds, enough time appears to have been available to develop a low-cost route while still successfully satisfying the mission constraints.

The implication of this behavior is that as some level of time becomes available in time-constrained problems, humans may induce errors over automation because the human falsely believes there is enough time to make an improvement, when in fact there is not. A more time-pressured condition would instead compel subjects to remain safe in replanning the initial route with small and local route modifications. Again, more study into these effects is warranted, including additional automation to aid the human in determining what types of modifications or decisions should be pursued.

In conclusion, the experimental findings show that replanning performance indeed varies with both automation type and time pressure. This alludes to the possibility of using some form of adaptive automation when integrating information for decision-assistance in time-critical and complex tasks. Adaptive automation would respond to the demands of the external environment and to the real-time performance of its individual user. In the increasingly complex and lethal nature of combat environments, the need is evident for in-flight replanning cockpit technology that can accurately and quickly assist pilots in making time-critical and life-dependent decisions. A user-centered approach to the design of decision-support automation is crucial for the successful implementation of replanner technology and other automated systems for cognitive assistance.

Acknowledgments

This work was supported by the Office of Naval Research under Award N00014-00-1-0659. We extend appreciation to Dr. Alan Natapoff, who was instrumental in the experimental design and post-experimental analysis of data. The views expressed in this paper are those of the authors and do not constitute the official policy or position of the USAF, DOD, or US Government.

References

Aust, R. 1996. Mission Planner Evaluation in the Mission Reconfigurable Cockpit. In *IEEE Conference Proceedings*, Vol. 1, A96-37576 10-1, 102-108.

Endsley, M. R., and Kiris, E. O. 1995. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, 37(2):381-394.

Layton, C., Smith, P. J., and McCoy, C. E. 1994. Design of a Cooperative Problem-Solving System for En-Route Flight Planning: An Empirical Evaluation. *Human Factors*, 36(1):94-119.

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*, 2000, 6(1):45-58.

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

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. 2000. A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 30(3):286-297.

Sarter, N. B., and Woods, D. D. 1995. How in the World Did We Ever Get into that Mode? Mode Error and Awareness in Supervisory Control. *Human Factors*, 37 (1):5-19.

Scerbo, M. W. 1996. Theoretical Perspectives on Adaptive Automation. *Automation and Human Performance: Theory and Applications*, R. Parasuraman and M. Mouloua, Eds. Lawrence Erlbaum Associates: Mahawah, NJ, 37-64.

Schulte, A., Stutz, P., and Klockner, W. 1999. Cognitive Concepts for Tactical Mission Management. In *Proceeding from the NATO System Concepts and Integration Panel Symposium*, Florence, Italy.

Sexton, G.A. 1988. Cockpit-Crew Systems Design and Integration. *Human Factors in Aviation.* E. L. Wiener and D. C. Nagel, Eds.. Academic Press: NY, 495-526.

Taylor, R. M. and Reising, J. 1998. The Human-Electronic Crew: Human-Computer Collaborative Teamworking. NASA CP 19990018356.

Wiener, E. L. 1988. Cockpit Automation. *Human Factors in Aviation*, E. L. Wiener and D. C. Nagel, Eds. Academic Press: NY, 433-461.