The Application of a Qualitative Model of Human-Interaction with Automation: Effects of Unreliable Automation on Performance

Scott M. Galster, Robert S. Bolia
Air Force Research Laboratory
2255 H Street
WPAFB, OH 45433-7022
USA

Raja Parasuraman
The Catholic University of America
250 O'Boyle Hall
Washington, DC 20064
USA

SUMMARY

A visual search paradigm was used to examine the effects of information automation and decision-aiding automation in a target detection and processing task. Manual, information automation, and decision-aiding automation conditions were manipulated with the size of the distractor set. Participants were required to respond to the presence or absence of a target in a time-limited trial. Reliability level (90%, 70%, 50%) of the automation was manipulated as a between subjects variable. Each reliability level group was comprised of eight volunteers for a total of 24 participants. Results indicated that the information automation cue condition engendered an increase in correct responses and a reduction in search times, regardless of set size or automation reliability. On the other hand, the presence of a decision-aiding cue differentially affected performance on all dependent measures as a function of both set size and automation reliability, alone or in concert with an information automation cue.

INTRODUCTION

The complexity involved in the acquisition, integration, and decision-aiding of many modern human-machine systems is such that performance efficiency can be compromised without the use of automation to support the operators of these systems. In response, real-time decision aids have been developed for a number of problem domains in both civilian and military applications. Unfortunately, automated aids have not always enhanced system performance, primarily due to problems in their use by human operators or to unanticipated interactions with other sub-systems. Investigations of human interaction with automation have revealed that automation does not always function in the way intended by designers and, moreover, can produce deleterious performance effects (Bainbridge, 1983; Parasuraman & Byrne, 2003; Parasuraman & Riley, 1997; Sarter & Woods, 1995; Wiener & Curry, 1980; Woods, 1996). Automation can change the nature of the demands on the operator and produce subsequent changes in performance not seen when automation is absent. Problems in human-automation interaction have included unbalanced workload, reduced system awareness, decision bias, mistrust, over-reliance, complacency, and reduced manual skills (Parasuraman & Riley, 1997).

Under conditions of high workload, time pressure, and uncertainty, these problems could become unwieldy and/or extract a high cost when present. For example, the potential for high fratricide rates in combat has led to the development of automated aids such as the Battlefield Combat Identification System (BCIS), which was designed as a decision aid for the identification of friendly troops by armor gunners. This system sends a microwave signal to interrogate a potential target and identifies it as friendly or unknown. The BCIS system

Paper presented at the RTO HFM Symposium on "The Role of Humans in Intelligent and Automated Systems", held in Warsaw, Poland, 7-9 October 2002, and published in RTO-MP-088.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
00 OCT 2003	**N/A**	**-**

4. TITLE AND SUBTITLE
The Application of a Qualitative Model of Human-Interaction with Automation: Effects of Unreliable Automation on Performance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory 2255 H Street WPAFB, OH 45433-7022; The Catholic University of America 250 OBoyle Hall Washington, DC 20064

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001577., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**10**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

was designed to improve target identification performance and reduce fratricide (Doton, 1996). As with many such automated aids, however, it is not clear whether performance with the system is in fact significantly improved, because of the aforementioned problems of human-automation interaction (Dzindolet, Pierce, Pomranky, Peterson, & Beck, 2001).

Similarly, the Theater High Altitude Area Defense (THAAD) system is a complex weapons system designed to intercept enemy short and medium range ballistic missiles. After enemy information is input from the field soldier the THAAD system makes recommendations on engagement decisions. These automated recommendations are then presented to the soldiers for their approval. The result of following an errant or unreliable recommendation could have serious consequences on the outcome of a mission. There is thus a pressing need for designing automation that supports military operators in command and control in ways that avoid such negative influences.

To minimize these costs of automation, several models have been proposed that have addressed the allocation of functions in automated systems. (see Fitts, 1951; Hancock, Chignell, & Lowenthal, 1985; Sheridan, 1980; Endsley & Kaber, 1999). Parasuraman, Sheridan, and Wickens (2000) have proposed a comprehensive model for the types and levels of human interaction with automation. Their model is based on a four-stage simplified human information-processing model. Parasuraman *et al*. (2000) use the information-processing model as an outline to describe a model of automated system functions. In their model, automation tasks can be assigned to a stage, pursuant to the function that it performs in the system. Accordingly, they name their stages information acquisition (acquisition), information analysis (analysis), decision and action selection (decision), and action implementation (action) respectively. Each stage comprises a unique continuum that will allow an expression of the varying degrees of automation.

To date, few studies have looked at the effect automation reliability levels have at different stages in overall system performance. In a study on automated cueing, Wickens, Conejo, & Gempler (1999) found that pilot detection performance decreased when a cue incorrectly guided attention away from the target even when the pilots knew the cue was not totally reliable. In another cueing study, Yeh, Wickens, & Seagull (1999) found that operators did not effectively pay attention to un-cued areas of a display. Other studies have found that while the automated aid has demonstrated a performance benefit, undesirable effects such as over-reliance in the automation (Galster, Bolia, Roe, & Parasuraman, 2001; Metzger & Parasuraman, 2001), poorer recall and automation-induced complacency (Horrey & Wickens, 2001) have also been found.

In an attempt to look at differential performance effects by stage, Crocoll and Coury (1990) examined decision-aiding performance when operators were given status, recommendation, or status *and* recommendation cues in an aircraft identification task. When the automation was perfectly reliable, performance across all levels of automation increased compared to the non-automated control group. However, when the automation was unreliable, detection performance for groups that received a recommendation, alone or in conjunction with status information, were markedly worse than for those who received the status only information. Subsequent studies (Sarter & Schroeder, 2001; Rovira, Zinni & Parasuraman, in press) have found similar results in an aircraft de-icing decision support system and in environments where the operator is engaged in multiple tasks.

The present study serves as a continuation of a series of planned comparisons between levels and stages of automation and the effect reliability levels have on action implementation. The first study (Galster *et al*., 2001) compared target detection in a manual and an automated information status cue in a basic visual search task. A visual search task was chosen as a simple simulation of a target identification environment, as used in BCIS or other automated systems for identification of friendly and enemy targets. The results indicated that there was a performance benefit attained with the presence of the automated aid, and that this benefit increased with the number of distractors. Moreover, these results were obtained without a concomitant increase in subjective workload. However, performance suffered when the automated cue was unreliable in the highest distractor set size, indicating an over-reliance on automation, which is consistent with the results of Yeh *et al*. (1999), who found that target detection performance increased with valid cues but decreased with invalid cues.

The second study (Galster, Bolia, & Parasuraman, in press), based on the same target identification task, revealed that a similar performance benefit was achieved with the presence of the information automation status cue indicating only the location of the target. Furthermore, this benefit increased with additional distractors within the search field. These benefits were realized even though the participants were aware that the information automation cue was not perfectly reliable. The second study also contained a decision-aiding automation cue that suggested a possible action to the participant. This cue did not produce a performance benefit however over and above the manual un-aided condition except when it was combined with the information automation cue.

These effects indicate that automated information cueing improves target identification performance under high target density conditions. In addition, the results demonstrated response time gains with the presence of the information automation (IA) cue, by itself or in conjunction with the decision-aiding (DA) cue, indicating that location is the enduring variable in reducing response times and increasing correct detections in this task.

To date, studies looking for detection and/or performance differences by stage of automation have not utilized a common task environment. The present study utilized the same basic visual search task with a manual, automation information, and decision-aiding cueing. But, unlike the second study the present study examined the manipulation of the level of reliability as a between subjects variable. Thus, we can compare the results from the first two studies to the present study with more confidence than we could if different task environments were utilized.

METHODS

Participants

Fourteen males and ten females between the ages of 18 and 32 (M = 21.92) served as paid participants. All participants were right-handed and reported normal or corrected-to-normal vision.

Experimental Design

A mixed design was employed in which 4 Automation Conditions (Manual (M), Information Automation (IA), Decision-Aiding (DA), Co-Located (IA + DA)) were combined factorially with 3 Distractor Set Sizes (10, 20, 30) to serve as within-subjects variables. Automation Reliability (90%, 70%, 50%) was manipulated as a between-subject variable.

Apparatus and Procedures

A visual search paradigm, in which participants were required to search a visual display for the presence or absence of a pre-defined target (╥) among similar distractors (╢, ╨, ╟, ╫), was employed. The display field emulated an artificial horizon consisting of 60% ground and 40% sky. Targets appeared only in the ground portion of the display.

All trials began with the presentation of a black fixation circle for 250 ms at the center of the display, followed by an interval of 1s in which the display was blank except for the artificial horizon. This was followed by the presentation of the automation cue(s) on the artificial horizon that lasted 300ms. The IA cue (a red plus sign) was always located in the green target area while the DA cue ("fire" or "no fire") was located in the blue-sky portion unless co-located with the IA cue. The automated cue(s) were cleared for 500ms and the target and distractor items were presented for 2.5s, or until the participant initiated a response. Trials were separated by an inter-trial interval of 2 s. A target was present on 50% of the trials. Participants were required to respond, using the left or right-arrow keys, to the presence or absence of the target, respectively. Each participant completed eight sessions of 150 trials during data collection. There were an equal number of trials in each session representing each of the three distractor set sizes. The trials were randomized with respect to both the number of distractors and the presence or absence of a target. The sessions consisted of two manual conditions, one of each IA, DA, IA+DA conditions of the prescribed reliability levels and on of each IA, DA, IA + DA conditions where the automation was 100% accurate. The conditions were counterbalanced with respect to the automation condition and the level of reliability experienced first.

All participants achieved a 75% correct response criterion in practice trials in each automation condition (under the 10 distractor condition) before experimental data collection began.

RESULTS

Correct Responses

A correct response was defined as the outcome of a trial on which a participant either correctly detected the presence of a high-priority target – indicated by the initiation of a "fire" response – or correctly judged the absence of a high-priority target – specified by a "no fire" response. Mean percentages of correct responses were submitted to a 4 (Automation Condition) × 3 (Set Size) × 3 (Reliability) repeated measures analysis of variance (ANOVA), revealing significant main effects of both Automation Condition, $F(3, 63) = 30.78$, $p < .05$, and Set Size, $F(2, 42) = 59.74$, $p < .05$, an Automation Condition × Set Size interaction, $F(6, 126) = 15.38$, $p < .05$, and an Automation Condition × Automation Reliability interaction, $F(6, 63) = 15.14$, $p < .05$. Neither of the other interactions was a significant source of variance. The Automation Condition × Set Size and Automation Condition × Automation Reliability interactions are depicted in Figures 1 and 2, respectively.

Inspection of Figure 1 reveals differences in search performance as a function of automation condition. Specifically, participants made more correct responses under the IA condition than under any of the other automation conditions or the manual control. Performance under the DA automation condition was only marginally different from performance under the manual condition for any set size. Further, correct response performance appears relatively stagnant under the combined IA + DA Automation condition across the three set sizes.

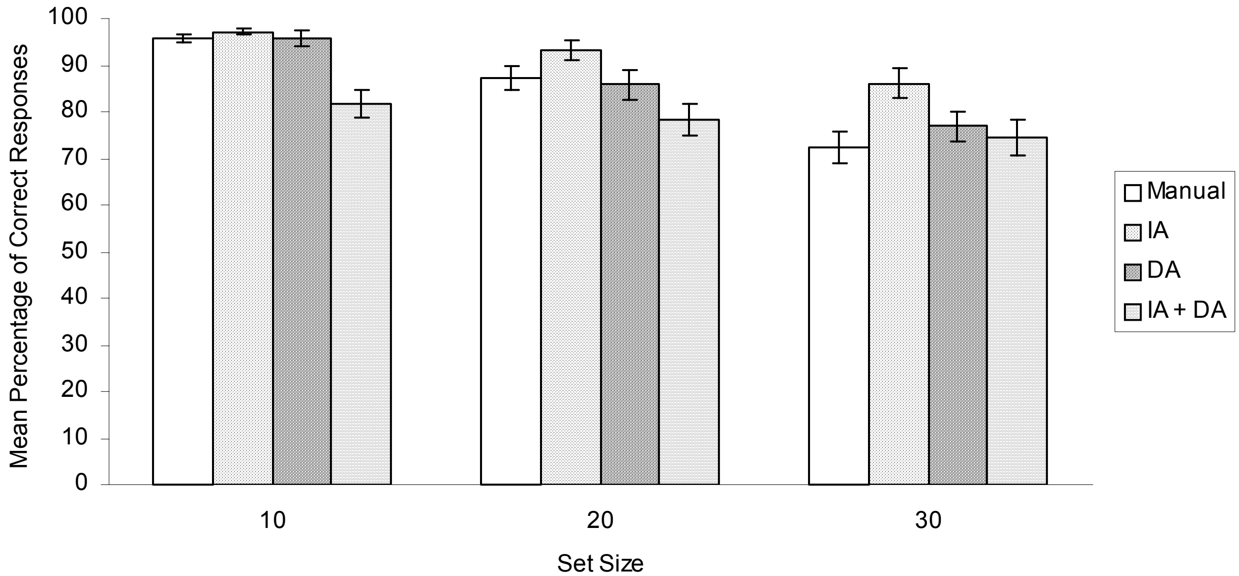

Figure 1: Mean Percentages of Correct Responses by Automation Condition as a Function of Set Size. Error bars represent one standard error of the mean in each direction.

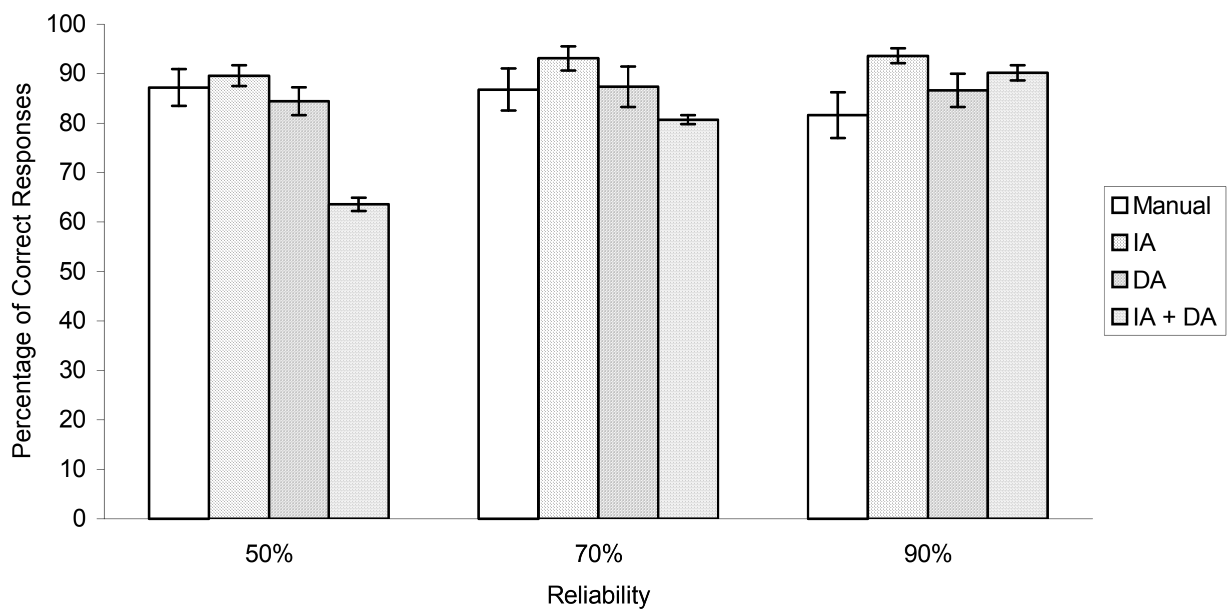

Figure 2: Mean Percentages of Correct Responses by Automation Condition as a Function of Automation Reliability Level. Error bars represent one standard error of the mean in each direction.

Figure 2 illustrates the interaction between automation reliability and automation condition. A major source of this interaction appears to be the increase in correct responses associated with increased automation reliability.

Figure 2 also suggests that the combination of information automation and decision aiding cues can lead to automation-induced complacency, a reduction in decision accuracy that can be caused by over-reliance in the automation. This effect is especially evident in the 50% and 70% reliability conditions.

For purposes of comparison, Figure 3 depicts the percentage of correct responses, as a function of whether the automation was not present (manual), always reliable (100%), or sometimes reliable (50%, 70%, or 90%), for each of the reliability groups, collapsed across automation conditions and set sizes. Thus, for example, the leftmost group of bars represents performance by participants in the 50% group under conditions of 100% automation, 50% automation, or no automation. These data show that, overall, manual performance and performance in the 100% automation reliability condition did not differ between groups. However, differences between the sometimes-reliable automation conditions are evident, with target acquisition performance increasing with increasing reliability.

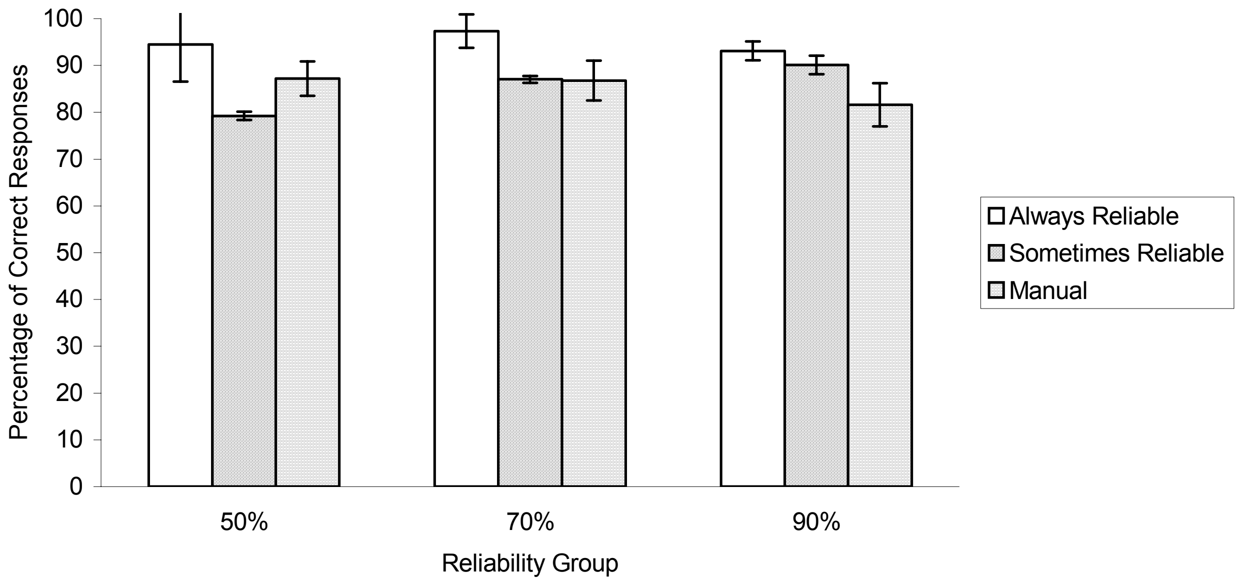

Figure 3: Mean Percentages of Correct Responses as a Function of Reliability Group and Level of Reliability. Within each group, the automation was either always reliable (100%), sometimes reliable (50%, 70%, or 90%), or unautomated (manual). Error bars represent one standard error of the mean in each direction.

Search Times

Mean search times of correct responses were submitted to an ANOVA analogous to that conducted for the percentages of correct responses. This analysis revealed significant main effects of both Automation Condition, $F(3, 63) = 37.85$, $p < .05$, and Set Size, $F(2, 42) = 85.55$, $p < .05$, an Automation Condition × Set Size interaction, $F(6, 126) = 5.32$, $p < .05$, and an Automation Condition × Automation Reliability interaction, $F(6, 63) = 8.22$, $p < .05$. None of the other sources of variances was significant. The Automation Condition × Set Size and Automation Condition × Automation Reliability interactions are presented in Figures 4 and 5.

The results depicted in Figure 4 are consistent with those obtained in previous studies in this series (Galster, Bolia, Roe, & Parasuraman, 2001; Galster, Bolia, & Parasuraman, *in press*). Namely, search times were reduced when an IA cue was present either alone or in conjunction with a DA cue. This effect is exacerbated under higher set sizes. This IA dominance effect is also visible in Figure 5, which demonstrates a decrease in search time with increasing automation reliability. This suggests that target acquisition and action implementation in a saturated complex visual field is enhanced most effectively by the presence of a reliable IA cue providing location information.

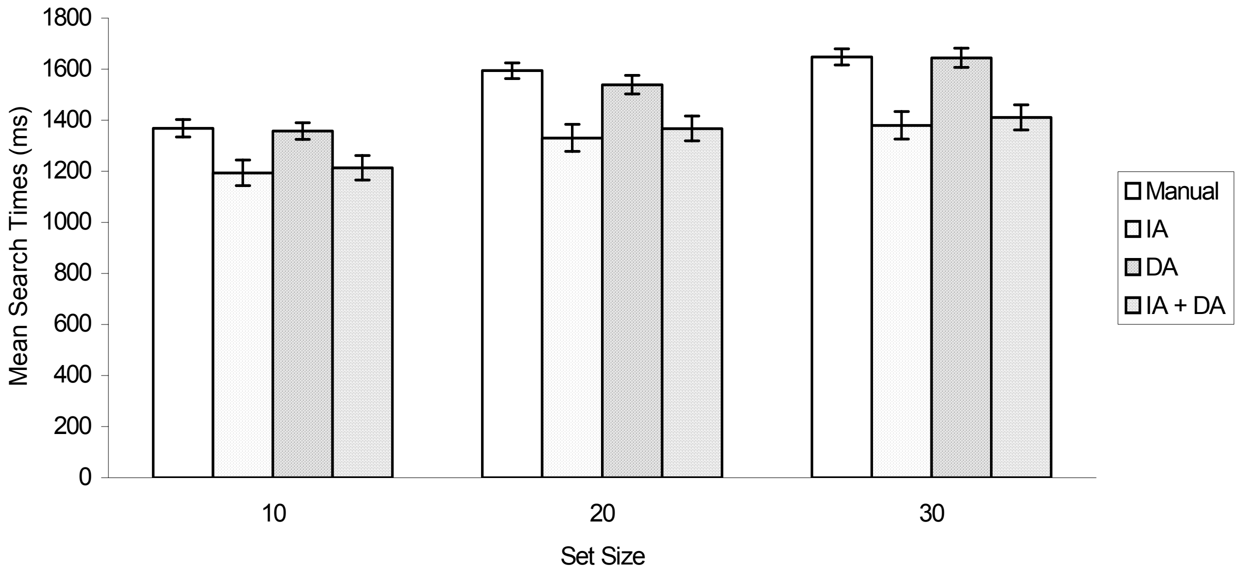

Figure 4: Mean Search Time (ms) as a Function of Set Size and Automation Condition. Error bars represent one standard error of the mean in each direction.

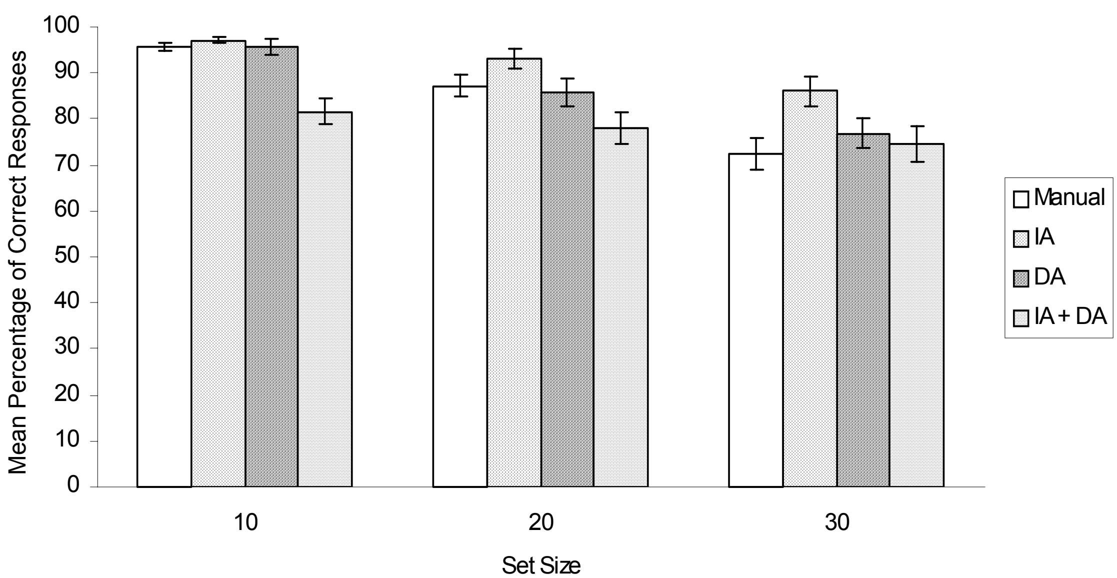

Figure 5: Mean Search Time (ms) as a Function of Automation Reliability and Automation Condition. Error bars represent one standard error of the mean in each direction.

Timeouts

Another source of variance not accounted for in the examination of correct responses is the number of trials in which a response is not made in the prescribed 2500ms allowed, termed a timeout. Mean percentages of total trials that resulted timeouts submitted to an ANOVA analogous to that conducted for the previous two analyses. This analysis revealed significant main effects of both Automation Condition, $F(3, 63) = 13.16$, $p < .05$, and Set Size, $F(2, 42) = 41.35$, $p < .05$, an Automation Condition × Set Size interaction, $F(6, 126) = 11.36$, $p < .05$, and an Automation Condition × Automation Reliability interaction, $F(6, 63) = 2.28$, $p < .05$. None of the other sources of variances was significant. The Automation Condition × Set Size and interaction is presented in Figure 6. Timeouts clearly increase with increases in set size, as expected. Of particular interest is the reduction of timeouts when the information automation cue was presented alone or combined with the decision-automation cue as compared with the other two automation conditions. This analysis is useful in determining that not only did the IA present sets of automation conditions reduce search times they also reduced the number of trials that ended due to a lack of a response.

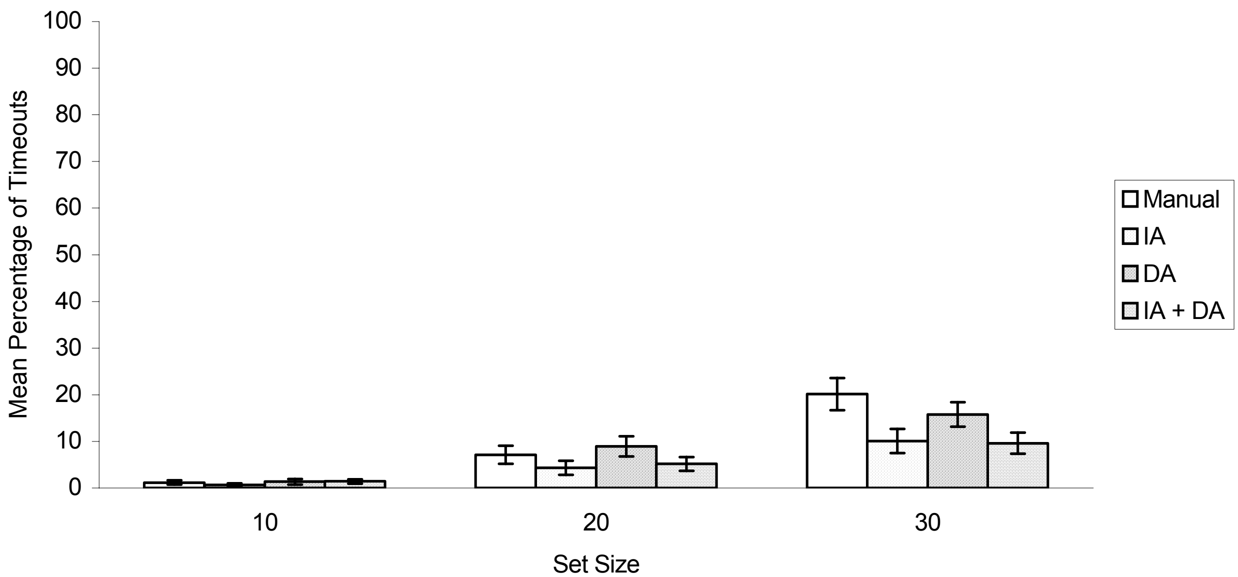

Figure 6: Percentage of Timeouts as a Function of Set Size and Automation Condition. Error bars represent one standard error of the mean in each direction.

DISCUSSION

The results of this experiment indicate that a performance benefit was achieved with the presence of the IA status cue. Furthermore, this benefit increased with the addition of more distractors within the search field and persisted under all reliability levels. These benefits were realized even though the participants were aware that the automated IA cue was not perfectly reliable.

The results also revealed that a performance decrement was present in the IA+DA condition for both set size and differences in the reliability of the automation. Of particular interest is the decrement found with this condition under the 50% reliability rate. This is most likely due to over-reliance on the automation to give the correct guidance resulting in an automation induced complacency effect under those conditions.

These effects indicate that automated information cueing improves target identification performance under high target density conditions. Thus the benefit of real-world battlefield or air defense identification systems might best be realized in complex, dense engagements, when the operator is likely to be already near their peak level of workload.

In addition, these results demonstrate response time gains with the presence of the IA cue, by itself or in conjunction with the DA cue, indicating that location is the enduring variable in reducing response times in this task. This effect was most prominent in the 90% reliability group but decreased as the reliability rate decreased.

This was the third in a series of studies looking at human interaction with automation by the stage it is presented. Comparative studies in different task domains should yield information regarding similar costs and benefits of automation by the stage it is introduced. Preparations have begun to determine if the present results are congruent in a complex and dynamic combat flight task utilizing the Synthesized Immersion Research Environment (SIRE) located at Wright-Patterson AFB near Dayton, OH.

REFERENCES

Bainbridge, L. (1983). Ironies of Automation. *Automatica,19*, pp.775-779.

Crocoll, W.M., & Coury, B.G. (1990). Status or Recommendation: Selecting the Type of Information for Decision Aiding. In *Proceedings of the Human Factors and Ergonomics Society 34th Annual Meeting*. Santa Monica, CA: HFES. pp. 1524-1528.

Dzindolet, M.T., Pierce, L., Pomranky, R., Peterson, S. & Beck, H. (2001). Automation Reliance on a Combat Identification System. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*. Santa Monica, CA. HFES. pp. 532-536.

Doton, L. (1996). Integrating Technology to Reduce Fratricide. *Acquisition Review Quarterly,* (*Winter Issue)*, 1-18.

Endsley, M.R., & Kaber, D.B. (1999). Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task. *Ergonomics, 42*, 462-492.

Fitts, P.M. (Ed.). (1951). Human Engineering for an Effective Air Navigation and Traffic Control System. Washington, DC: National Research Council.

Galster, S.M., Bolia, R.S, & Parasuraman, R. (in press). Effects of Information Automation and Decision-Aiding Cueing on Action Implementation in a Visual Search Task. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Santa Monica, CA. HFES.

Galster, S.M., Bolia, R.S., Roe, M.M., & Parasuraman, R. (2001). Effects of Automated Cueing on Decision Implementation in a Visual Search Task. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*. Santa Monica, CA. HFES. pp. 321-325.

Hancock, P.A., Chignell, M.H., & Lowenthal, A. (1985). An Adaptive Human-Machine System. In *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, (18), pp. 627-629.

Horrey, W.J., & Wickens, C.D. (2001). Supporting Situation Assessment through Attention Guidance: A Cost-Benefit and Depth of Processing Analysis. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*. Santa Monica, CA. HFES. pp. 316-320.

Metzger, U., & Parasuraman, R. (2001). Conflict Detection Aids for Air Traffic Controllers in Free Flight: Effects of Reliable and Failure Modes on Performance and Eye Movements. In *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.

Parasuraman, R., & Byrne, E.A. (2003). Automation and Human Performance in Aviation. In P.S. Tsang and M.A. Vidulich, *Principles and Practice of Aviation Psychology* (pp. 311-356). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Parasuraman, R., Sheridan, T.B., & 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*, 286-297.

Rovira, E., Zinni, M., & Parasuraman, R. (in press). Effects of Information and Decision Automation on Multi-Task Performance. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*. Santa Monica, CA. HFES.

Sarter, N., & Schroeder, B.K. (2001). Supporting Decision-Making and Action Selection under Time Pressure and Uncertainty: The Case of In-Flight Icing. *Human Factors, 43 (4)*, 573-583.

Sarter, N., & 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*, 5-19.

Sheridan, T.B. (1980). Computer Control and Human Alienation. *Technology Review, 83*, Oct., pp. 60-73.

Treisman, A. & Gelade, G. (1980). A Feature Integration Theory of Attention. *Cognitive Psychology*, 12, 97-136.

Wickens, C.D., Conejo, R., & Gempler, K. (1999). Unreliable Automated Attention Cueing for Air-Ground Targeting and Traffic Maneuvering. In *Proceedings of the Human Factors and Ergonomics Society the 34th Annual Meeting*. Santa Monica, CA: HFES. pp. 21-25.

Wiener, E.L., & Curry, R.E. (1980). Flight-Deck Automation: Promises and Problems. *Ergonomics, 23*, 995-1011.

Woods, D.D. (1996). Decomposing Automation: Apparent Simplicity, Real Complexity. In R. Parasuraman and M. Mouloua (Eds.) *Automation and Human Performance: Theory and Applications* (pp. 1-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Yeh, M., Wickens, C.D., & Seagull, F.J. (1999). Target Cueing in Visual Search: The Effects of Conformality and Display Location on the Allocation of Visual Attention. *Human Factors*, *41*, 524-542.