THE RISING COGNITIVE COST OF AUTOMATION

MACINNES, W. JOSEPH
Faculty of Computer Science/Department of Psychology, Dalhousie University,
Halifax, N.S. Canada. B3H 4J1
Email macinnwj@cs.dal.ca
Phone (902)-494-7025

MCCABE, JOHN
School of Health and Human Performance
Dalhousie University,
Halifax, N.S. Canada. B3H 4J1

KLEIN, RAYMOND
Department of Psychology, Dalhousie University,
Halifax, N.S. Canada. B3H 4J1

ABSTRACT

Recent studies have examined the trade-off involved in auto-pilot systems. Although automation reduces the workload of the pilot in important tasks, it also reduces the 'situation awareness' and produces 'Out of the loop unfamiliarity' (OOTLUF). Whereas the workload reduction provides a benefit, the loss of situation awareness entails a cost when the pilot/performer resumes control. This present study explored this trade-off by asking participants to navigate a set of virtual Hebb/Williams mazes, under both manual and automated conditions. Automation was simulated on alternate mazes by having participants learn a route through which they were passively transported using a previous participant's active exploration. The cost of situation awareness was measured as the increase in time to complete a 3-Dimensional (3D) maze based on the method of learning, while workload was measured using a sensory task incorporated into the maze navigation. Although the 'OOTLUF' cost of automation was observed in this experiment, the benefits to workload were not as clear cut.

KEY WORDS
Automation, Workload, Situation Awareness, Virtual, Maze.

INTRODUCTION

Automation is becoming more and more prevalent as technology advances. With systems becoming more complex and the desire to optimize performance getting greater, automation is taking on an ever increasing role in our society. The main reasons for automation include economic benefit (assembly lines), taking over tasks which a human operator is incapable of performing (launching rockets), as well as to reduce the workload of a human operator (auto-pilot). It is this final reason that requires the most interaction with human operators and is the focus of this study.

Recent studies (3, 1) have shown that although automated systems elicit a benefit in the workload of the operator, they also seem to come with a cost. Nearly all studies report a reduction in awareness of the operator with the surroundings and the state of the system. This reduction in situation awareness has been labeled 'Out-of-the-Loop-Un-Familiarity' (OOTLUF) or Pilot-out-of the-Loop (POOTL)(3). When an operator is in control of a system, there is a closed loop of instruction, system response and feedback that keeps the operator aware of the current state of the system. During automation, the operator is removed from this loop and may lose the significance of any feedback, or not get any feedback at all. A real world example of this can be seen in car drivers versus car passengers: Many people report being able to learn a route faster while driving than being a passenger.

This tradeoff between workload reduction and loss of situation awareness has been the focus of many studies which have attempted to maximize the benefits while minimizing the costs of automation. At the heart of these studies is the issue of how to measure workload and awareness. One approach has been to administer surveys such as the NASA-TLX (Task Load Index) (5) for workload and the Situation Awareness Global Assessment Technique (SAGAT) (4) for situation awareness. Although these self report methods introduce a component of participant bias, proponents of these methods claim that outcome measures such as reaction time tasks are too disruptive to the main task in the experiment.

The purpose of this study is to use a real-time virtual environment to measure the trade-off between situation awareness and workload that arises from automating a system. To minimize participant bias, outcome measures were used to measure both variables. The overall time to complete the 3D virtual maze was used as the measure for situation awareness and reaction time to unpredictable probes was used as a measure of workload. There were two factors that had to be considered in deciding an appropriate probe; The probe could not be internal to the maze or the measure would be affected by situation awareness, but the probe could not be completely independent of the original task or it would be too disruptive. The solution was to present a probe that looked like a net blocking progress through the maze. Participants were told before the experiment that this net represented them being 'captured' during their attempt to complete the maze and they could only release the net by pressing the space bar as fast as possible. This allowed for a probe that was outside of the maze itself, yet tied to the task objective to minimize disruption.

METHODS

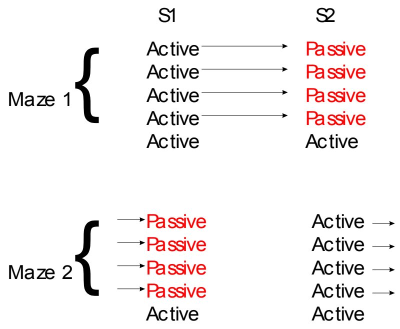
Thirty-Six adult volunteers (19 Female, 17 Male) participated in the experiment. All participants were either students or employees of the university.

Scenes were rendered on a Pentium 120 running Windows 95 and using the OpenGL Graphics library (Version 1.1). No 3D acceleration was used and rendering averaged at 10 frames per second. The display was presented on a 17" (Viewsonic) monitor, with 16 bit color at 120Hz. Rendering was paused during probe presentation to prevent interference with response timing. Input was accepted through the keyboard, with the 4 arrow keys controlling navigation, and the space bar used for probe response.

Environments were designed as virtual duplicates (10) of the standard Hebb/Williams mazes (6,8). Mazes Six unique practice mazes and twelve unique test mazes were used. Mazes were measured in abstract units, six square units with a one square unit alcove at both the starting and finishing corner. The starting alcove had a red roof and the finishing alcove a green roof in order to differentiate the two. Navigation through these mazes used a fluid motion, first person perspective and was set to a velocity one units per second and a turn rate of 50 degrees per second. It was possible, using multiple arrow keys, to both move and turn simultaneously. Movement, controls, and collision detection were implemented to mimic popular 3D video games (eg. Quake, Unreal) in order to use experience in these games as a variable.

Participants completed a set of practice mazes until they were able to complete a set of nine in under 168 seconds. This time was selected based on previous studies' (10) time of 158 seconds plus an additional ten seconds for probe acquisition.

Image 1



Participants completed twelve unique mazes in the main portion of the experiment with each maze occurring five times in sequence. A blank maze (empty room with clear line of sight to the goal) was provided between each set of five to provide a break between mazes. All participants, on alternating sets of mazes, were latched to the performance of previous participants for the first four of the five runs through that maze (See figure 1). Due to previous studies showing gender differences in spatial tasks (10), participants only viewed the performance of the same gender. One male and one female participant navigated all mazes actively to act as 'seeds' for the remaining participants. The data for these two participants were not included in the analysis.

Probes were presented randomly and frequently throughout the maze. There was roughly a 10% cumulative chance of a probe occurring every second up to a maximum of four probes per maze in Experiment 1 (E1). This was increased to roughly 14% per second up to a maximum of ten probes in Experiment 2 (E2).

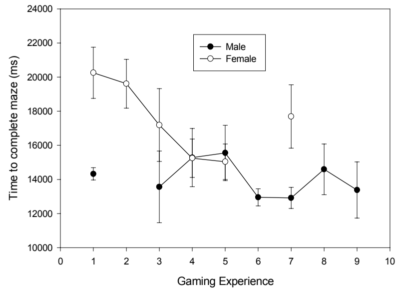
Participants were asked to rate their computer expertise before the start of the experiment. Computer literacy and computer gaming experience were rated on a scale from one to ten with one being no knowledge at all and ten being expert. This was done to determine if experience in similar tasks had an effect on performance.

RESULTS

situation awareness

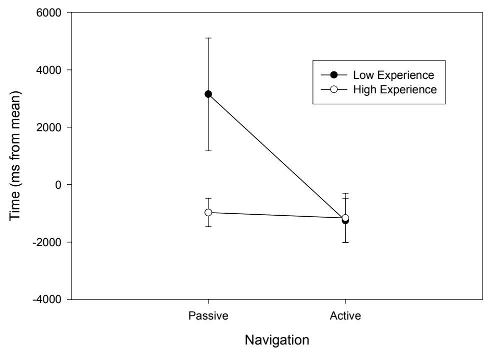
Only the time to complete the last of each set of five mazes was used in the analysis of situation awareness because all participants actively navigated the fifth maze. As expected, there was a significant effect of both maze and gender measured by the time to complete the maze. Later mazes tended to be more difficult than earlier mazes(P<0.001) and men tended to complete the mazes faster than women (P<0.001). Since these results were expected and not the focus of this study, they were controlled for in all further analysis of awareness. This was done by normalizing all maze times to become the difference between the actual time and the mean time for that gender and maze. It is worth noting, however, that it was not possible to determine if this gender effect was a result of previous gaming experience since the two variables strongly co-varied (see figure 2). Males tended to report having the greatest amount of experience in 3D computer games while females tended to have the least. This may support recent studies that suggest gender differences in spatial tasks might be learned (9).

Figure 2



The new normalized scores were analyzed using a 2 (Automation) x 2 (Gaming Experience) repeated measures, analysis of variance (ANOVA). Computer knowledge and gaming experience were rated as 'high' or 'low' based on whether they were above or below the mean response for that question. Automation was based on whether the participant passively or actively navigated the previous four attempts at that maze.

Figure 3



There was a significant effect of automation (F(1,32)=6.2,P<.02) Reflecting a 2.4 second cost on the fifth maze when the first four were passively navigated. There was also an interaction between gaming experience and automation (F(1,32)=5.3,P<.03) reflecting that the measured cost of automation was much greater for participants with low ratings for gaming experience (See figure 3).

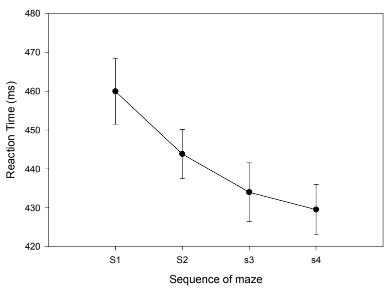
workload

Probe reaction times were subjected to non-recursive, sliding criterion, outlier analysis program (11). The resulting data was analyzed using a 4 (Maze Sequence) x 2 (Gender) x 2 (Gaming experience) x 2 (Automation) repeated measures ANOVA. Maze sequence refers to the order that a participant navigated the first four of each five maze series. Only the first 4 mazes were used since the fifth maze was always navigated actively and could not be split by Automation. Automation in this analysis refers to whether the *current* maze was being passively navigated (Whereas automation in the previous analysis refers to the method of learning in the *previous* four mazes). Again, there was an effect of Gender ($F(1,30)=7.2, p<0.02$) with men being faster than women. While this re-enforces the idea that men tend to have a lower workload in these tasks (whether due to experience or gender), it is not the focus of this study and therefore not used in further analysis. There was a significant main effect of sequence ($F(3,96)=14.7, p<0.0001$) with participants responding progressively faster throughout the sequence. Although there was no main effect of automation ($F(1,32)=0.8, p=.37$), it did interact with sequence ($F(3,96)=2.8, p<.04$) as well as marginally with gaming experience ($F(1,32)=4.0, p=.06$).

DISCUSSION

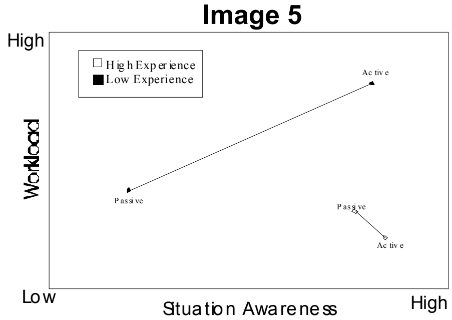
The typical results found in studying the effects of automation show a benefit to the workload of the operator, yet a cost in the situation awareness. Although the results of this study demonstrate the expected cost in terms of lowered situation awareness, the results do not seem to demonstrate the expected benefit to workload. Possible explanations may include: a) the probes used in the experiment were not capable of measuring workload. b) The benefits to workload were being offset by something akin to low vigilance. c) Some, or all, of the participants were already at high/optimal efficiency and could not benefit from workload reduction.

Figure 4



The possibility that the probes were not an accurate measure of workload seems unlikely due to the main effect of sequence. As participants progressed through multiple iterations of the same maze, we would expect a reduction in workload as well as a corresponding reduction in probe reaction time, and this is exactly the result we see (figure 4). Low vigilance in passive navigation was a concern from the start, and the reason for the random frequent probes. To further test this possibility, both frequency and the maximum number of probes were increased in E2 to determine the effect on probe reaction time. If low vigilance was masking the benefits to workload, this manipulation should at the least show a benefit to probe reaction times. There was, however, no significant effect of increasing the probe frequency ($F<1$).

Image 5



The most likely possibility is that workload is already low or optimized and cannot benefit from the automation. A closer look at the interaction with computer experience, shows that participants with low computer experience do show a benefit of automation while those with high computer experience actually show a slight cost (See figure 5).

Showing the tradeoff as a graph (See figure 6), participants with low experience show the typical tradeoff involved in automation, where the highly experienced participants do not. In fact, since the lower right side of the graph is the optimal position (High awareness, low workload), experienced participants are actually the most efficient during non-automated, active navigation.

CONCLUSION

Although automation is capable of producing marked benefits in a number of areas, it is a mistake to assume that these benefits occur across all conditions. While studies have shown that the costs of automation can be manipulated and even optimized, it is important to realize the effects of these manipulations on the benefits of automation. In particular, experienced operators seem to react to automation differently than novices. If the prime objective of the automated task is to reduce the workload of the operator, it is important to ensure that the operator is not already working at optimal workload.

REFERENCES

(1)Endsley, Mica R. and Kaber, David B. (1999) Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42, 462 - 492.

(3)Endsley, M.R. and Kiris, E.O. (1995) The Out-of -the-loop performance problem and level of control in automation. *Human Factors*, 37, 381 - 394.

(4)Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D.J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: LEA.

(5)Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research, in P.A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam:Elsevier Science/North Hollond), 139 - 183;

(6)Hebb, D. O., & Williams, K. A. (1946) A method of rating animal intelligence. *Journal of General Psychology*, **34**, 59-65.

(8)Rabinovitch, M. S., & Rosvold, H. E. (1951) A closed field intelligence test for rats. *Canadian Journal of Psychology*, **5**, 122-128.

(9)Sedighian, K. & Klawe, M. (1996) Super Tangrams: A Child-Centered Approach to Designing a Computer Supported Mathematics Learning Environment. ICLS.

(10)Shore, D., Stanford, L., MacInnes, J., Klein, R. & Brown, R., (2001). Of Mice and Men: Using Virtual Hebb-Williams Mazes to Compare Learning across Gender and Species. *Cognitive, Affective, & Behavioral Neuroscience*, 1 (1), 83-89.

(11)Van Selst, M and Jolicoeur, P. (1994) A Solution to the Effect of Sample Size on Outlier Elimination. *The Quarterly Journal of Experimental Psychology, 47a, 631 - 650.*