



**VARIABLE TIMING EFFECTS ON PERFORMANCE AND BEHAVIOR  
WITHIN HUMAN-MACHINE TEAMS**

JUNE 2017

Mark A. Harris, Capt, USAF

AFIT-ENV-MS-17-J-056

**DEPARTMENT OF THE AIR FORCE  
AIR UNIVERSITY**

**AIR FORCE INSTITUTE OF TECHNOLOGY**

---

---

**Wright-Patterson Air Force Base, Ohio**

**DISTRIBUTION STATEMENT A.**  
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENV-MS-17-J-056

VARIABLE TIMING EFFECTS ON PERFORMANCE AND BEHAVIOR WITHIN  
HUMAN-MACHINE TEAMS

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the  
Degree of Master of Science in Systems Engineering

Mark A. Harris, BSEE

Captain, USAF

June 2017

**DISTRIBUTION STATEMENT A.**  
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-17-J-056

VARIABLE TIMING EFFECTS ON PERFORMANCE AND BEHAVIOR WITHIN  
HUMAN-MACHINE TEAMS

Mark A. Harris, BSEE

Capt, USAF

Committee Membership:

Michael E. Miller, DAF, PhD  
Chair

Clay Koschnick, Lt Col, USAF, Ph.D.  
Member

Christina F. Rusnock, Maj, USAF, PhD  
Member

### **Abstract**

When well designed human-agent teaming is used within a system, it provides a tremendous benefit than neither the human or automation can accomplish alone. This research sought to explore various parameters that would help in the effective design of robust adaptive automation. Agent timing and its interaction with spawn rate were explored from a performance perspective and to understand the number of times a human engages with the environment. Surprisingly, agent timing or certain user actions within the environment did not significantly affect performance but significant changes in workload were observed. Adaptive automation design seeking to maximize human-agent team performance should have a thorough understanding of the human experimentation results needed to explore the effect of an artificial agent's timing on the performance of a human-agent team within a highly dynamic task environment. The research contained within explores those human experimentation results.

## **Acknowledgments**

I would like to express my sincere appreciation to my faculty advisor, Dr. Michael E. Miller, for his guidance and support throughout the course of this thesis effort. The insight and experience was immensely appreciated. I would, also, like to thank Shannon Enix, for her support and constant encouragement provided to me in this arduous endeavor.

Mark A. Harris, Capt, USAF

## Table of Contents

	Page
Abstract .....	iv
Table of Contents .....	vi
List of Figures .....	viii
List of Tables .....	ix
I. Introduction .....	1
General Issue .....	1
Problem Statement.....	2
Research Objectives .....	3
Research Focus .....	4
Investigative Question .....	4
Methodology.....	5
Assumptions and Limitations .....	8
II. Literature Review .....	10
Chapter Overview.....	10
Automation System Design.....	10
Relevant Research .....	16
III. Methodology .....	20
Chapter Overview.....	20
Experimental Environment.....	20
Participants .....	22
Experimental Design and Procedure .....	23
Data Analysis.....	25
IV. Analysis and Results.....	28

Performance as Measured by Score .....	28
Workload .....	30
Human Draws .....	32
Variable Agent Delay with Dynamic Spawn Rate .....	34
Desirability of Control and Score .....	34
Human Draws and Score .....	37
Discussion.....	37
V. Conclusions and Recommendations .....	41
Investigative Questions Answered .....	41
Conclusions of Research .....	42
Recommendations for Future Research.....	43
Significance of Research .....	44
Appendix A - Game Play Pre-Questionnaire.....	46
Appendix B – Desirability of Control.....	47
Appendix C – Game Play Post-Questionnaire.....	48
Bibliography .....	50

## List of Figures

	Page
Figure 1: Space Navigator Game.....	7
Figure 3: Spawn, Agent Delay and Normalized Score.....	30
Figure 4: Spawn, Agent Delay, Workload Analysis.....	32
Figure 5: Spawn Rate, Agent Delay, and Human Draws Analysis .....	33

## List of Tables

	Page
Table 1: Data Collected and Data Analyzed.....	24
Table 2: Maximum Score Computations .....	26
Table 3: Desirability of Control Measures.....	36

# VARIABLE TIMING EFFECTS ON PERFORMANCE AND BEHAVIOR WITHIN HUMAN-MACHINE TEAMS

## I. Introduction

### General Issue

Adaptive Automation has been proposed to address the deficiencies that are associated with human interaction with automation (Endsley M. , 1987; Wiener, 1988; Parasuraman R. , 1993; Endsley & Kiris, 1995; Parasuraman & Riley, Humans and automation: use, misuse, disuse and abuse, 1997). The earliest approach to automated systems design was to automate anything that was possible to be automated without consideration for the operator in the system (Kaber & Endsley, 2003). The integration of the human and the machine was an afterthought. Thus, this approach led to the operator having to devise the best strategy to be successful with the machine, which would often mean circumventing automation aids. The result of this design approach did not always lend itself to creating a more effective human-machine team and, in many cases, reduced both human performance and the resulting system performance. For example, by automating tasks, which reduce the operator's knowledge of the state of the system, Out-of-the-loop (OOTL) situation awareness decrements can occur (Wickens & Wickens, 1982).

With the application of human-centered design to automated systems, the operator's performance can be considered during the design of the system to improve the human-machine system and lead to better performance. By applying the human-centered design approach, the design focus shifts from the unpolished process of automating

anything that is possible to designing with the focus on the operator's performance and situation awareness (SA). The later approach provides a new adaptive automation design framework referred to as human-centered automation (Billings C. , 1991). The purpose of human-centered automation is to design systems that keep the human operator engaged in control loops which may require operator involvement with meaningful and well-designed tasks that operators are able to perform. The goal of such human-centered automation is to optimize the performance of overall human-machine team (Billings C. , 1997). Currently, the field of automation is embracing human-centered automation, but there is still very little research detailing the exact process of how to explicitly design adaptive automation within a system (Steinhauser, 2009). This research will examine a few design parameters of an adaptive automation system to better categorize what constitutes effective design. Specifically, the variability surrounding agent delay time, desirability of control, and the optimization of human effort with automation will be assessed.

### **Problem Statement**

Adaptive automation can be extremely useful for operators. It can increase the task load that an operator can comfortably handle, improve situation awareness by freeing the operator from menial tasks so they can concentrate on the larger picture, increase the safety of a system, and increase the likelihood of mission success (Parasuraman R. , 1993; Parasuraman R. M., 1996; Hilburn, Molloy, Wong, & Parasuraman, 1997). However, for all of the positive potential benefits, there is a lack of

sufficient literature describing a process to design an adaptive automation system into either existing or new systems.

Recent research has included the proposal for a design process model for considering human-automation allocation (Bindewald, Miller, & Peterson, 2014). One particular subset of the problem is the optimal time that automation should engage to assist the team (Goodman, Miller, & Rusnock, 2015). This research has suggested that the optimal time for automation engagement would likely change as a function of task load (Goodman T. , Miller, Rusnock, & Bindewald, 2017). For example, as the rate of environmental events requiring action decreases, it is hypothesized that the human will be more likely to have time to respond to these events and will therefore likely perform better with a less aggressive automation.

### **Research Objectives**

The current research will explore the use of a variable automation assistance to explore the interaction of optimal automation assistance time with task load in a highly dynamic environment. This automation assistance will scan the environment continually, and after a predetermined amount of time, if no action is taken will assist the user in performing the task. It is believed that through the use of variable parameters, system designs can be identified which ensure that the operator will be able to successfully meet the full range of system and task load demand that can occur across a range of expected missions while maintaining engagement. The research objective will be to investigate six proposed hypotheses related to the manipulation of automation assistance, task load,

desirability of control, and workload optimization with human-machine teaming performance.

### **Research Focus**

This research aims to expand the body of knowledge on the design of adaptive automation and increase the understanding of how varying the time in which an automation agent initiates, after a system defined agent delay, will affect a human-machine team's performance. Furthermore, this research aims to provide analysis for desirability of control, a user specific attribute that may affect the effectiveness of a human-machine team.

### **Investigative Question**

The overarching research question is “How can one apply knowledge of the variability within a highly dynamic environment to aid the design of adaptive automation systems?” This research will examine six hypotheses.

1. The first hypothesis states that the rate of automation assistance will have a significant impact on the performance of the human-machine team and the relationship will have a negative correlation.
2. The second hypothesis asserts that the workload, as experienced by the participant, will be negatively correlated to the rate of automation assistance. Essentially, the increased assistance provided by the automation will be correlated to lower subjective workload scores.

3. The third hypothesis is that human engagement in the environment will be higher during a low rate of automation assistance. The reasoning for this, is that as the automation teams with the user and assumes more responsibility, the human's workload is lessened and therefore their actions on the environment will decrease as the automation takes a more active role.
4. The fourth hypothesis is that as the task load is decreased the rate of automation assistance should also be reduced. This change would allow the operator to become more engaged as the task load is reduced.
5. The fifth hypothesis proposed that the participant's desirability of control scale measure would be inversely related to their performance. That is individuals who seek to maintain more control will rely less on the automation, resulting in lower performance, especially under high task load conditions.
6. The last hypothesis predicts that the operator's behavior in the environment would be positively correlated to the performance of the human-machine team. This means that as the operator becomes more engaged in the system, the performance of the overall team would increase. The assumption embedded within this hypothesis is that, given sufficient time, the human will outperform the automation. Therefore, as the operator takes more of the task load of the human-machine team by taking more actions the environment, the overall performance will be increased.

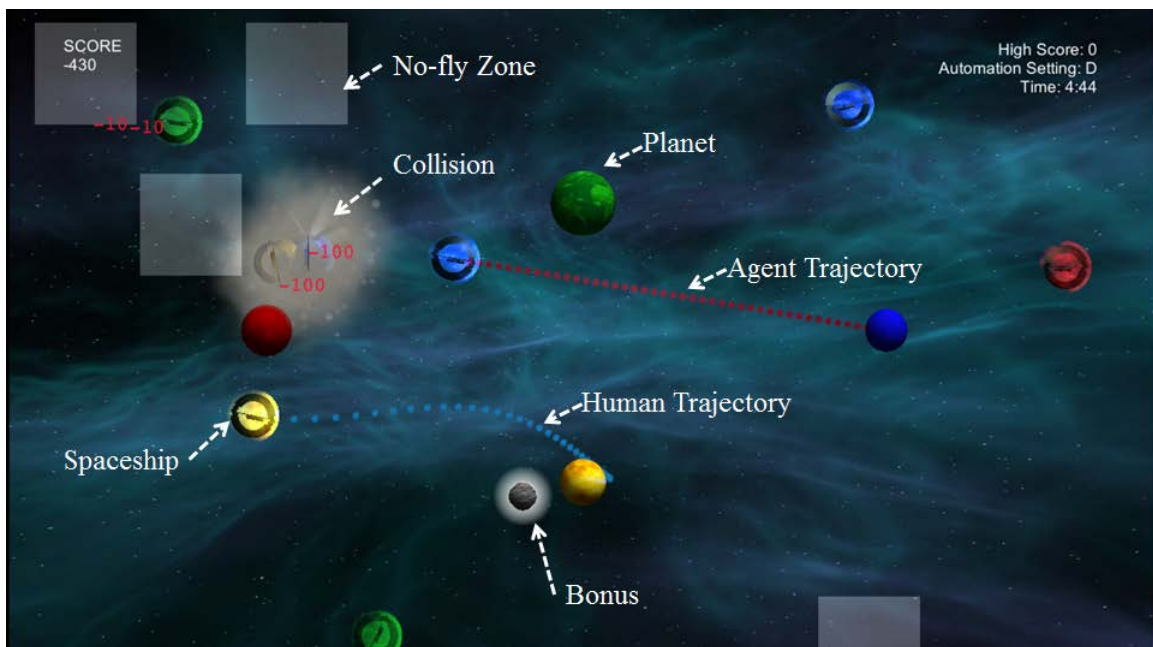
## **Methodology**

This experiment will use an air traffic control style game, called Space Navigator shown in Figure 1 as the highly dynamic environment to explore relationships between

operator performance, task load, and behavior. The game begins with spaceships being spawned anywhere along the edges of the screen with an initial trajectory that the operator must select and draw a path from the spaceship to its corresponding destination planet, delineated by color (e.g., an operator must draw a trajectory from a red space ship to the red planet). Simultaneously, the operator drawing the path from the spawned spaceship to its corresponding destination planet must avoid obstacles (other space ships and randomly moving no fly zones) and acquire point bonuses in the form of randomly appearing orb-shaped objects. Task load will be manipulated within this environment by changing the rate at which these ships are spawned.

The game consists of a five-minute session, in which the goal is to score as many points as possible. The player can earn positive points in one of two ways: landing a spaceship on its destination planet and “picking up” a randomly appearing bonus (the ships path crosses over the bonus location and is automatically scored as a bonus). The player or operator loses points in two ways: permitting two ships to collide, thereby destroying both ships, and permitting a spaceship to enter a no-fly zone. A no-fly zone is represented by a rectangle that moves at fixed intervals and to random locations. Points are lost for every second the ship remains in the no-fly zone and presents an additional task load for the human operator. Ships appear at predetermined intervals along the edge of the screen and have an initial trajectory into the gameplay area. There are many things that change the player’s workload in the game. Specifically, the player’s workload varies with changes in the rate of appearance of spaceships, the length of their trajectories, the rate of appearance of bonuses and the rate at which no-fly zones move. The player’s perceived workload can also be changed by the variable agent delay. In this game the

agent scans the screen and draws a route for any untasked ships after a certain amount of time, where this time is referred to as agent delay time and is used to represent the rate of automation assistance. This automation agent assists the human-machine team by drawing a straight line from the ship to its corresponding color destination planet. The agent delay times for this experiment are 2.6, 5.6, and 8.6 seconds. The number and location of planets are fixed throughout the game.



**Figure 1: Space Navigator Game**

Embedded within the Space Navigator game is a data collection system, which permits an experimenter to gather a wealth of information. Every event created by the system (spawning spacecraft, bonus locations, no fly zones, spawn rate, collision, agent draw, etc) and those events caused by the human operator (trajectory path drawn, bonuses collected, ship landed safely, collisions, trajectory path redrawn, etc) are captured from

the game. Using data from both the system and the human operator's interaction with the system will provide insight into how the operator performs tasks with manual control and as a member of a human-machine team. Data such as the average time that a ship spends in transit from being spawned to its destination planet, or the number of times that a ship has been tasked with a trajectory are examples of the types of human behavioral information provided by the Space Navigator data tracking system. Game score represents human performance.

### **Assumptions and Limitations**

Space Navigator represents a highly dynamic environment which will be used to test broad hypotheses with regard to human-machine performance, operator workload and behavior. Within the framework of the Space Navigator experiment, human-machine performance will be measured by the normalized score. The workload was captured by NASA-TLX questionnaires that each operator reported at the end of each session. The human behavior characteristics that were of importance in this research were related to human draws.

This experiment used 10 participants, that each completed two practice rounds and ten five-minute trials of the Space Navigator with automation game. The most apparent limitation is the small number of participants used. A more robust experiment would include a larger number of participants. Additionally, another limitation is the need to eliminate 20% of the data points in order to calculate a two-way repeated measures ANOVA. This experiment collected data across three different spawn rates (2, 3, and 4 seconds) and three different agent delay times (2.6, 5.6, and 8.6 seconds) however in

order to run statistical tests, such as the two-way repeated measures ANOVA, only the four furthest edges of the data set were used which eliminated 20% of the data points. While the total number of samples was well over 100 distinct points, more data would increase robustness of analysis.

At the beginning of this research, the assumption was made that the operators would be performing very close to the optimal arousal and performance with help of the high rate of automation assistance. This relationship between stress and performance is known as the Yerkes-Dodson Law. This law, originally developed by psychologists in 1908, says that performance increases with physiological or mental arousal, but only up to a certain point. When those arousal levels become too high, the performance decreases. The process is often described as an upside “U” shape with the arousal of the participant as the independent variable and the performance measure from weak to strong as the dependent variable. All of the assumptions made throughout this paper, assume that the operators are in the sweet spot of arousal and high performance when the rate of automation assistance is high (i.e., 2.6 seconds) and the task load is high (i.e, spawn rate of 2s) (Goodman T. , Miller, Rusnock, & Bindewald, 2017). This experiment assumes that a high rate of automation assistance will keep the operators adequately engaged but not overwhelmed in the dynamic environment. However, decreases in task load might require decreases in the rate of automation assistance to keep the operator at peak performance.

## **II. Literature Review**

### **Chapter Overview**

The purpose of this chapter is to review existing literature on adaptive automation, specifically the human-centered design of systems employing automation. Specifically, this chapter will examine the present design rationale for human-centered automation, review literature which examines the effectiveness of this rationale, and provide a discussion of the use of workload in system design and evaluation.

### **Automation System Design**

Numerous problems have been documented which have led to human-machine system errors due to the improper design of automation systems (Wiener and Curry 1980, Moray 1986, Billings 1991, Sarter and Woods 1995). These problems are linked to issues that arise when human operator capabilities and limitations are not appropriately considered; problems include vigilance decrements, complacency, and loss of situation awareness (SA). These problems often lead the human operator to be out-of-the-loop (OOTL) and therefore unable to respond appropriately during a critical incident. OOTL problems steal valuable human performance from the overall human system team (Young, 1969; Kessel & Wickens, 1982).

Out-of-the-loop performance problems are characterized by a decreased ability of the human operator to intervene in system control loops and assume manual control when needed while supervising automated systems. These problems occur when the human operator places too much reliance on automation (Itoh, 2011), places undue trust in the automated system (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003), or loses

situation awareness preventing a timely response during an automation system failure (Itoh, 2011). It is for these noted reasons that the relationship between SA and workload creates a conflict for those designing automation. When attempting to optimize both SA and workload within the context of an automation program, the designer runs into a conundrum. Under high-levels of automation the operator will be under tasked and will experience a low workload condition, which causes the operators to experience boredom and fatigue due to the lack of mental involvement. The resulting boredom and fatigue will cause the operator to sacrifice SA as the user disengages and fails to project potential future circumstances (Endsley & Kiris, 1995; Endsley & Kaber, 1999; Pope, Comstock, Bartolome, Bogart, & Burdette, 1994). Under these conditions, operators will find it difficult to reorient themselves during times of system failure or unpredicted events, resulting in out-of-the-loop performance problems.

Conversely, under high task loads, users will experience excessive workload due to a lack of automation, therefore cognitive overload may occur. High workload can also lead to the operator having low SA and task performance as the operator becomes overwhelmed with providing a timely response to the system. As the operator is over tasked, this would prevent them from projecting upcoming responses, also reducing the operator's ability to respond to system failures or unpredicted events. Furthermore, the increasing task requirements beyond which the operator can effectively manage can lead to feelings of frustration and defeat which contributes to a loss of SA, thereby decreasing human-machine system performance (Kaber, Riley, Tan, & Endsley, 2001). The challenge of adaptive automation is to identify and keep the system within the optimal workload or functional range that will allow for the best level of operator SA, so that total

system performance will be optimized. The early design method of automating anything that can be automated will not suffice for effective adaptive automation design.

Optimized system operation requires a human-centered design approach.

Human operators, acting as supervisors or monitors, have a hard time detecting system errors and performing tasks manually in the event of an automation system failure (Wiener and Curry 1980, Billings 1988, Wickens 1992). Additionally, humans have a hard time just monitoring a system. Billings (1988) reviewed six major aircraft accidents and was able to trace the failure directly to the monitoring of automated systems or the flight parameters controlled by the automated systems. When operators are not directly involved in the operation of the system, they are less likely to detect critical system states that would precede major system failures. Therefore, passive monitoring will hamper the effective recovery of a system from a major system failure and can have fatal consequences in many different environments ranging from nuclear plants to commercial and military flight.

In addition to being ineffective at detecting errors and poor system states when they occur, human monitors require significantly more time to reorient themselves to the current state of the system and choose the appropriate response. Wickens and Kessel (1979, 1981) showed through lab demonstrations that longer system recovery times and poor response accuracies occur for operators who had been removed from the control loops before a critical event occurred that necessitated intervention. The two types of situation awareness (SA) problems which lead to an increased incidence of failure that can occur in automated systems with a monitor have been attributed to the following three mechanisms:

- (1) Decreased change in vigilance and complacency associated with monitoring;
- (2) The operator taking a passive role as monitor instead of an active role in controlling the system; and
- (3) Changes in the quality or form of feedback given to the human operator.

To overcome the shortfalls of technology-centered design methods to automation, a more human focused design methodology has been adapted called, human-centered automation. Human-centered automation is a phrase adapted to describe human-centered design when applied to systems employing human interaction with automation (Billings C. , 1997). Furthermore, the phrase expresses the need for automation to work in conjunction with the humans operating or interacting with it; rather than creating the automation and expecting the human operator to adjust to the design of the automation (Sheridan & Parasuraman, 2006). In his work, Billings (Billings C. , 1997) pushes the discussion forward by laying out criteria for the design of human-centered automation. Researchers like Bindewald and Goodman go on to expand on Billings early efforts to produce and design an automation design approach that seeks to maximize performance through automation while optimizing user engagement (Bindewald, Peterson, & Miller, 2014).

According to Billings, (1997) the goal of human-centered automation is to create systems that keep the human in the control loop with meaningful and well-designed tasks that operators are capable of performing to optimize overall human-machine system function. Billings further explains that human-centered automation should ensure that

automation does not leave the human with a fragmented and difficult job. It should define the assignment of tasks to a human and computer in controlling an automated system such that a team effort is achieved (Billings C. , 1997; Endsley M. R., 1996).

Sheridan proposed that human-centered automation had many alternative meanings from ‘allocate to the human the tasks best suited to the human, allocate to the automation the tasks best suited for it’ through ‘achieve the best combination of human and automatic control’, where “best” is defined by explicit objectives (Sheridan T. , 1997). The definitions set forth by Sheridan span from a function-oriented perspective to a mission-oriented perspective. Recent research by Bindewald has used the definition of adaptive automation that defines it as the dynamic allocation of levels of automation or other aspects of a system (Bindewald, Peterson, & Miller, *Creating Effective Automations that Maintain Explicit User Engagement*, 2014).

An approach to human-centered automation that attempts to balance machine and human task allocation deal with, “How much level of automation (control) should be used?” The other question asks “Which tasks should or should not be given to the automation agent for execution?” One approach seeks to optimize the assignment of control between the human and automated system by keeping both involved in system operations. This has been called Level of Automation (LOA) or ‘level of control’ (Draper, 1995). The other method acknowledges that control must pass back and forth between the human and automation over time, depending upon the context of the situation and seeks to find ways of exploiting this relationship to increase human performance. This has been called Adaptive Automation (AA) or Dynamic Function Allocation (DFA) (Corso and Moloney 1996).

Sheridan (1997) presents various degrees of automation that were defined in terms of the autonomy of complex system information sensing and control execution. These degrees are correlated to the extent to which the human or machine was involved in the system control and the level of aid the automation was providing to the operator. The LOA approach defines the allocation of system control between a human and computer in terms of the degree to which both are involved in system operations (Endsley, 1996). The objective of this approach is to find levels of automation that are optimal for human capabilities and capacities during a specific situational context.

Through human-centered automation, high levels of human-machine system performance may be achieved that ensures that the human has the capability to monitor the system, that they receive adequate feedback in the state of the system, and that the automation functions in predictable ways (Billings 1997). Each of these support achievement of SA. Furthermore, Parasuraman et al. (1992) said that AA represents an optimal integration of automation with the human operator based on the level of operator workload. Parasuraman et. al placed high importance on the state of the operator in determining the optimum level of integration. Scerbo (1996) surmised that under AA, different types of automation may be started and terminated dynamically based on the situational demands placed on the system, inclusive of the operator. Kaber and Riley (1999) contributed a modern definition of AA, saying that it concerns the scheduling of the allocation of control between a human operator and computer over time, with the intent of improving human performance as part of complex system operations or maintaining operator involvement in system control to reduce OOTL performance problems.

## **Relevant Research**

Research conducted by Endsley and Kiris (1995) involved a study of simulated automobile navigation task and found that SA was lowest under full automation and only slightly lower under the intermediate LOAs as compared to the fully manual task. Additionally, Endsley and Kiris found that lower SA correlated with OOTL performance decrements when the automated aid failed and operators were forced to intervene and maintain manual control for the duration of the automation failure. Of the three sources of OOTL performance decrements involved passive processing was the only culprit in the study. Results showed that intermediate LOAs had value in reducing the OOTL issues that could arise from full automation.

Endsley and Kaber delved deeper into this problem space within the framework of their 10-level LOA taxonomy (Endsley & Kaber, 1999). Their research used a dynamic control task, Multitask<sup>®</sup> (Endsley & Kiris, 1995), which is a simulation of a radar-monitoring task that an air traffic controller or air surveillance radar operator would perform. The simulation uses multiple targets which must be eliminated prior to reaching their expiration time or crashing into one another. Each target yields varying amounts of reward points for its successful elimination and penalty points for mission failures. The idea behind the Kaber and Endsley simulation was that multiple targets would compete for operator attention simultaneously forcing the operator to develop a strategy to maximize performance.

During the test trials, the operators would experience simulated automation failure, where the LOA would shift from their assigned LOA to complete manual control. The failure in the automation, forced the human operator into complete manual control of

the system and would occur at random time intervals and last a short duration. During the course of the experiment, various simulation freezes were instituted in order for the human operator to respond to a Situational Awareness and Global Assessment Technique (SAGAT) query to measure situational awareness, and a NASA-TLX was also collected to measure operator-perceived workload (Endsley & Kaber, 1999).

The results from their experiment showed that LOA significantly affected both task performance and OOTL performance problems. Several important findings on the topic of automation came from this particular Endsley and Kaber research. First, performance was improved by computer aiding in the range of LOAs from 2-10, however performance was hindered at the LOAs that involved joint human-computer option generation (LOAs 4 – 8 or Shared Control to Automated Decision Making). The automation that aided in the action selection aspect of the task (LOAs 6, 8, and 9 or Blended Decision Making, Automated Decision Making, or Supervisory Control) did not significantly affect system performance as compared to the baseline of automating human-decision making. Endsley and Kaber's research concluded that performance at high LOAs (9 and 10, or Supervisory Control and Full Automation) was better than Manual Control performance.

The experiment that provided the foundation for the topic of this thesis was conducted using 32 volunteers to play the Space Navigator game manually with varying types of automation assistance under relatively high workload conditions. Bindewald designed the agent to respond slightly faster than the average human response and to use similar, dissimilar, or straight line draws as the automation assistance (Bindewald J. M., 2015). The results of Bindewald's experiment showed that straight-line automation had a

statistically significant difference in the operator's reliance on the automation to create more complex initial trajectory draws. Additionally, the qualitative results showed that the humans preferred the straight-line assistance over other types of trajectory drawing algorithms because of its predictability.

The follow-on experiment sought to hone the time which the automation would engage itself (Goodman T. , Miller, Rusnock, & Bindewald, 2017). Using 20 participants, Goodman and colleagues set out to test different delay times and explain how adjusting agent delay timing, changes the task that the human and machine perform in the given environment. This experiment examined the effects of shifting the agent delay time from 0.1 seconds to 11.6 seconds. The results determined that human-machine performance improves as the automation activates more rapidly. Additionally, there is a slight change of trajectory draw length between 0.1, 2.6, and 5.6 seconds in the manual redraws of the automation drawn trajectories. Ultimately, this research concluded that the right balance of explicit engagement and performance by triggering the straight-line automation was around the 2.6 second agent delay time for the spawn rate explored in these experiments. Additionally, an important concept that surfaced from his results was that the addition of automation that can draw trajectories for the human changes the human's underlying task from one of "trajectory drawing" or "ship routing" to a completely different task of "trajectory correction" or "collision avoidance". This means that careful consideration should be taken to design automation to ensure that both human-machine performance is improved and user engagement is maintained.

Additionally, Goodman, developed a human performance and workload model to predict system performance and to provide further insight into the phenomena observed

in Bindewald's research. This model employed the data from the study conducted by Bindewald to obtain task time distributions and other behavioral and performance data for the human both in the absence and presence of the straight-line agent. This model predicted that a maximum score would be achieved if the human drew a higher percentage of trajectories when there were fewer ships on screen, which corresponded to lower workload conditions. Additionally, the model predicted that the human would draw a higher percentage of trajectories when a large number of ships were on screen, which likely corresponded to conditions where the agent was unable to draw trajectories that would not result in collisions.

### **III. Methodology**

#### **Chapter Overview**

This experiment explored the teaming of a human and automation agent within a representation of an air traffic control management system through a modified tablet game called *Space Navigator*. Specifically, this study is designed to answer the question, “How does the timing of automation assistance within a dynamic task load environment affect system performance and human behavior?” The human’s objective within the study was to task individual spaceships that appear on screen, by drawing a route from that particular ship to its corresponding planet color to achieve the maximum number of points. This methodology section will review the steps taken before, during, and after the experiment to collect data. A brief summary of the data collected and analysis performed will be presented along with the independent variables.

#### **Experimental Environment**

There are five objects that are acted upon throughout the game; spaceships, planets, routes, bonuses, and no-fly zones. Spaceships are entities that are “spawned” and “fly” onto the screen with a random trajectory, and must be acted upon by the user. The goal of each spaceship is to land on its matching color planet destination. The spaceships appear at two, three and four second intervals depending on the game play setting. Each spaceship is spawned and appears at a random location along the edge of the screen. In order to avoid collisions, consecutive spaceships cannot come from overlapping locations (i.e. they must be more than a spaceship’s width away from the location of the last ship’s appearance). A spaceship will appear on the screen moving toward the screen edge

opposite from which it was spawned. Planets are fixed destinations on the screen. Each spaceship is assigned a destination planet which matches the color of its corresponding spaceship. When the edge of a spaceship overlaps the edge of the destination planet, a spaceship lands on the planet and points are given to the player accordingly. Furthermore, spaceships cannot land on planets that are not their destination planet and will instead traverse through them. A route within the game is defined as a visible path of movement that a spaceship is assigned to follow from its current location to the end of the path drawn by the participant. A successful route, emanates from a spaceship and ends at the given spaceship's destination planet while unsuccessful routes will end in collisions or traversals off the screen. A successful route is drawn by touching the finger on top of the given spaceship, dragging the pressed finger across the screen, and releasing at the ship's destination planet. The path traced by the finger is recorded as the ship's route, and is displayed on the screen as the series of red dots. Additionally, the automation assistance in the form of trajectory draws will initiate a route on spaceships which have been spawned but not tasked within a predetermined amount of time. Finally, the ship will automatically traverse its completed route until it lands on a planet or reaches the end of the route, in which case it will continue in the direction of the last heading.

Bonuses are objects that can be picked up by spaceships to gain extra points. A bonus appears randomly at any location on the screen. They appear at time intervals according to a uniform probability distribution. A bonus is obtained if the ship's edge comes in contact with the bonus edge. Any ship can pick up any bonus. Upon a ship picking up a bonus, the player is allotted the number of points the bonus is worth; once the bonus appears it does not disappear until collected by a ship. No-fly zones are the

areas that randomly appear on the screen. Each no-fly zone is represented by a small rectangle that covers a portion of the screen. A spaceship that enters a no-fly zone will lose points according to the amount of time it is in the no-fly zone. Specifically, for each second in the no-fly zone, the player loses ten points. A spaceship is in a given no-fly zone when any part of the spaceship overlaps any part of the no-fly zone. The no-fly zones will move at set intervals to random places on the screen.

Points can be obtained in two ways: successfully landing a ship on a planet and picking up a bonus. Points can be lost in two ways: traversing “no-fly zones” and when two or more ships collide.

## **Participants**

The experiment involved 10 male volunteers with an average age of 27.1 years and a range of 23 to 34 years. The experiment involved the use of a computer based tablet game environment; therefore each participant was asked how often they use laptops, tablets, desktops, phones, and gaming consoles. On average, they used tablets roughly 1 – 2 times a month. Other computer based platforms, including smart phones, were reported being used 3 – 7 times a week. Of the 10 participants, 9 identified as right hand dominant and 1 identified as left hand dominant. The participants were screened for color blindness in previous experiments and would not have been allowed to participate if color blindness was detected.

## **Experimental Design and Procedure**

The experiment consists of three distinct phases. During phase 1 each participant was required to complete a pre-experiment interview to determine his or her exposure to systems of this type measure their desirability of control. The data from this was collected from the questionnaire as listed in Appendix A. Additionally, each participant was given a desirability of control questionnaire which is found in Appendix B. Finally, each participant completed two five-minute practice instances of *Space Navigator*. The participant was given a tablet computer with the *Space Navigator* loaded.. A complete game lasts five minutes and must be un-interrupted. Participants were advised on proper use conditions. Upon completion of the task, the participant would move on to the next experiment phase.

The purpose of phase 2 was to gather significant game-play information surrounding a participant's interaction with an automated agent that aids them in playing *Space Navigator*. Each participant is required to complete fifteen five-minute instances of *Space Navigator* with adaptive automation. Each participant will be presented three iterations of five agent delay and spawn time settings. The settings will include the independent variable of spawn times (time between ship arrivals on screen) of 2, 3, and 4 seconds and agent delay time of 2.6, 5.6, and 8.6 seconds. The dependent variable in this experimental design is performance as measured by overall game score. The sequence of the combination of the two variables were counter-balanced using a Greco-Latin Square Design. After each game, participants were given a NASA-TLX battery questions. Upon completion of the task, the participant returned the tablet computer and the recorded data was uploaded for processing. Data is anonymized through a number assignment.

The last and final phase was to gather data based on the entirety of the experiment to assess the person’s experiences with the adaptive automation system. Each participant was required to complete a brief interview. The interviews assessed the participants’ subjective views about the experiment and their performance in it. This qualitative information was not used in the experimental analysis presented herein and questions are found in Appendix C.

The data collected in this experiment consisted of nine total data points of which, only five were used as data points and four analyzed for statistical significance and other analysis. The nine data points were a mixture of the 2, 3, and 4 second spawn rates while the agent delay times consisted of the 2.6, 5.6 and 8.6 second times. Table 1 below shows the sum of all data points with information collected, the data of interest that was used as support throughout this research and finally the data that was statistically analyzed.

**Table 1: Data Collected and Data Analyzed**

Spawn Rate	4	X		X
	3		O	
	2	X		X
		2.6	5.6	8.6
Agent Delay Time				

The order that the participants ran through the sessions was counterbalanced. Counterbalancing is a method for controlling the order effects in a repeated measures experiment. When a counterbalanced experiment is run, it is control for order effects, in which a separate group of subjects, each receive treatments in a different order.

Counterbalancing, allows one to use a general technique to manipulate two or more independent variables within-subjects, without having each participant go through every possible combination and order of effects. This experiment used a Graeco-Latin square design in which every row and every column contains each combination of independent variables only once. This set up allowed for the counterbalancing of effect order.

### **Data Analysis**

The score analysis was completed by taking the average of scores per individual per condition. These scores were then normalized by dividing the score by the total number of possible points achievable. Increases in score may occur as a function of spawn rate for one of two reasons. First, and most interesting to the current research, the score may increase due to decreasing spawn rate because human-agent team performance might increase. Second, the score may increase simply because there are more ships available with which to score points as spawn rate decreases (i.e., there are twice as many ships available when a ship is spawned every 2 seconds as when ships are spawned every 4 seconds). This relationship is expressed in Table 2 which indicates the maximum score with and without the bonuses for each spawn rate. For this reason, it was necessary to normalize the score by the total number of possible points to isolate the team performance effects from the attributes of the game.

**Table 2: Maximum Score Computations**

<b>Spawn Rate (seconds)</b>	<b>Maximum Score w/o Bonuses</b>	<b>Maximum Score with Bonuses</b>
2	15000	15750
3	10000	10750
4	7500	8250

For the most accurate normalized score measure, the participants' scores were divided by the maximum score including bonuses of their corresponding condition and then multiplied by 100 to obtain their normalized score for each trial. Therefore, for further analysis, score will be analyzed using the normalized values.

A similar process was accomplished for the measures of the draws throughout this experiment. The measures of the automation draws, initial human draws, human redraws, the sum of all human draws, and the automation routes that were redrawn by the human were all normalized based upon the number of ships that were presented during the experiment. For the spawn rate of 2 seconds, there are a total of 150 ships that can be spawned. By normalizing the number of ships across each trial, the number of draws can also be normalized. Therefore, the number of draws was multiplied by 2 for spawn rates of 2 seconds and for the 4 second trials, the number of draws was multiplied by 4 to obtain the normalized draw value.

The average *Space Navigator* score and normalized score was then plotted with the standard error for each spawn rate and agent delay combination across all participants. The primary method of data analysis was performed using a two-way

repeated measures Analysis of Variance (ANOVA). Eta-squared ( $\eta^2$ ) was used to measure the statistical effect size of the sample means. The normalization was performed to allow each of the conditions to be compared in magnitude to the others because the number of ships and opportunity to score points, was not the same across all conditions.

The workload analysis was completed by taking the average of all NASA-TLX workload measures, per individual and condition. The average NASA-TLX workload score was then plotted with the standard error for each spawn rate and agent delay combination. A two-way repeated measures ANOVA, with effect size was also conducted on the workload measures and used to provide a measure of the standardized difference between the various sets of normalized scores means. Behavioral changes were investigated by analyzing the normalized number of human draws and agent draws. Finally, linear regression was used to measure the correlation between operator behavior and normalized score. The variables corresponding to human behavior for the regression analysis were desirability of control and normalized number of human draws for this portion of the analysis.

## IV. Analysis and Results

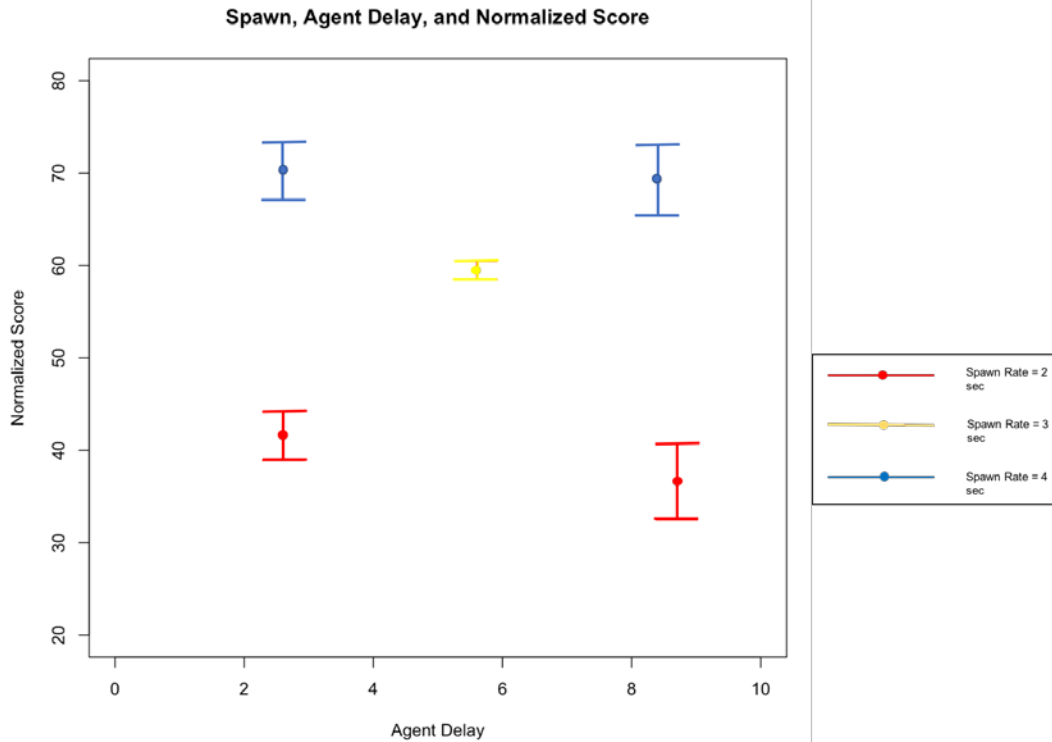
### Performance as Measured by Score

A two-factor, repeated measures ANOVA indicated that the effect of the spawn rate was significant for normalized score,  $F(1, 9)=159.8$ ,  $MSE=118$ ,  $\eta^2=0.63$ ,  $p>0.0001$ . An example of the code base for the ANOVA calculation is found in Appendix D. As the effect size is large (i.e.,  $\eta^2>0.5$ ), spawn rate has a very significant effect on normalized score, with normalized score increasing significantly as fewer ships are spawned. Furthermore, the ANOVA showed that the effect of the automation agent delay time on the user's final score was not statistically significant,  $F(1, 9)=3.55$ ,  $MSE=48.2$ ,  $\eta^2=0.0062$ ,  $p<0.092$ , therefore the first hypothesis presented is rejected. The agent delay time did not have an effect on the normalized score. Furthermore, the results of this experiment showed that, as the spawn rate decreased, the user did not perform better, on average, with the less aggressive automation settings. The human-machine team performed best when the spawn rate was low and the automation was high.

Possible reasons for these results could include; the sample size was not large enough (10 participants), or users that have a higher desirability of control could be less prone to effective human-machine teaming. Similar studies, conducted by Bindewald and Goodman have used total participants in the range of 20, 32, and 36 individuals (Bindewald J. M., 2015) (Bindewald, Peterson, & Miller, 2014) (Goodman, Miller, & Rusnock, 2016). As presented earlier, the effect of desirability of control and normalized score measures will be explored later as a potential effect on normalized score and human-machine teaming. Finally, the ANOVA indicates that the interaction between spawn rate and agent delay time did not have a statistically significant effect on the

normalized score  $F(1, 9)=1.02$ ,  $MSE=79.50$ ,  $\eta^2=0.0058$ ,  $p<0.339$ . As the effect size for both the agent delay time and the interaction of agent delay time with spawn rate is small, these effects are likely not of practical significance. Overall, the results show that spawn rate has an overwhelming effect on the normalized score measure. When the human is less tasked by the environment, they perform relatively better, regardless of agent delay time.

Figure 2: Spawn, Agent Delay and Normalized Score shows a plot of the normalized score as a function of the agent delay time. The results of the normalized score show that the spawn rate of 2 seconds produced consistently lower normalized scores than the 3 second and 4 second spawn rates. Furthermore, the mean of the normalized score for the 2.6 second agent delay times and 4 second spawn rate was slightly larger than the means of the normalized score for the 8.6 second agent delay time, although not significantly larger.



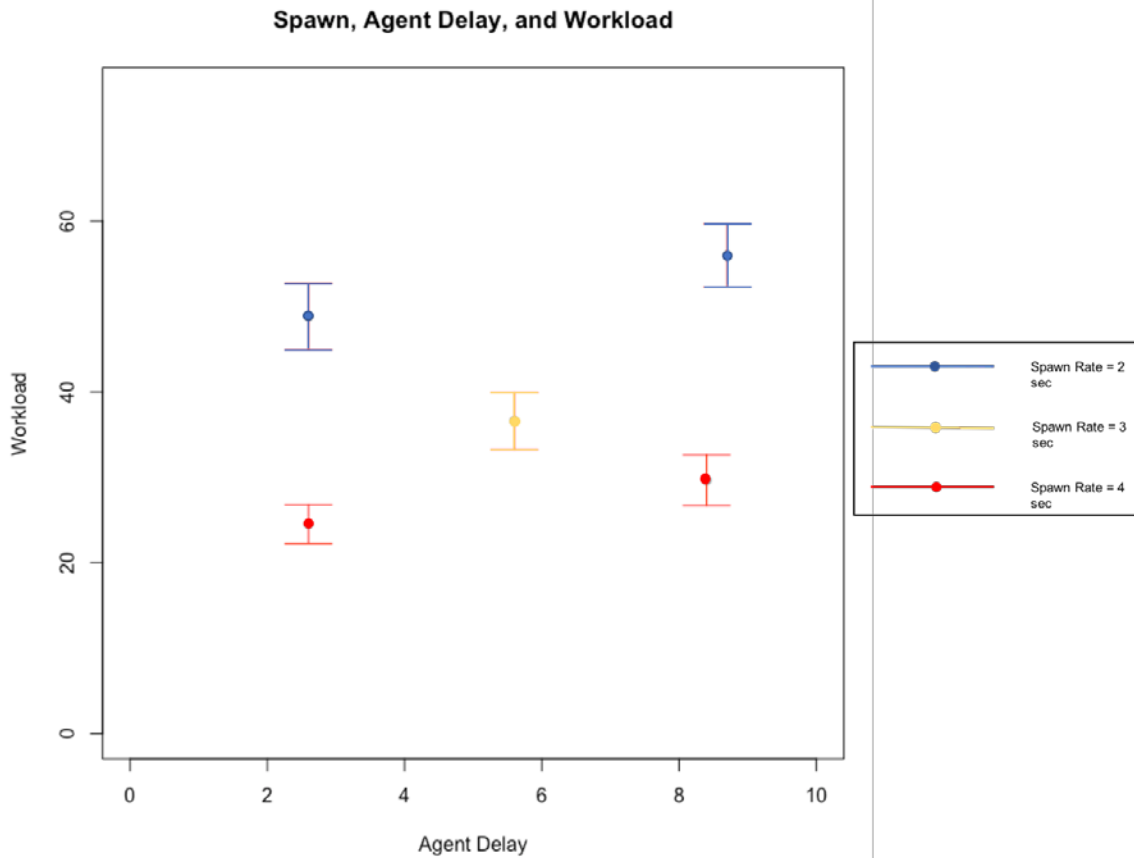
**Figure 2: Spawn, Agent Delay and Normalized Score**

## Workload

A two-factor, repeated measures ANOVA indicated that the effect of the spawn rate was significant for workload,  $F(1, 9)=42.79$ ,  $MSE=150$ ,  $\eta^2=0.601$ ,  $p>0.0001$ . The effect size of  $\eta^2$  indicates that spawn rate has a large effect on workload. This result emphasizes the previous conclusion that the spawn rate has a direct effect on how the user performs and perceives their workload within the trial. Furthermore, the ANOVA showed that automation agent delay time did have a significant effect on the user's perceived workload  $F(1, 9)=17.99$ ,  $MSE=21.1$ ,  $\eta^2=0.036$ ,  $p<0.0022$ . This analysis is important because it confirms the second hypothesis presented at the beginning of this research; the agent delay time has a statistically significant effect on the workload

measure. Finally, the two-factor repeated measures ANOVA indicates that the interaction between spawn rate and agent delay time does not have a statistically significant effect on the workload  $F(1, 9)=0.208$ ,  $MSE=48.15$ ,  $\eta^2=0.0027$ ,  $p<0.339$ .

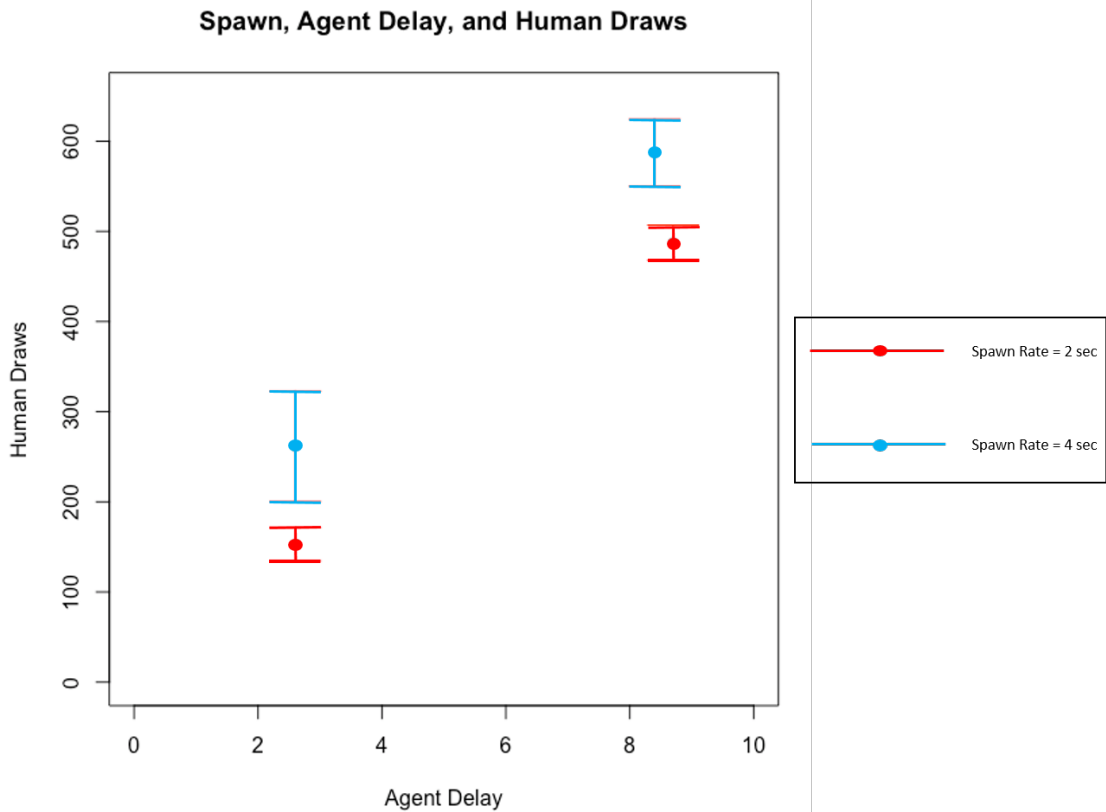
The workload analysis in Figure 3 was completed by taking the average of all NASA-TLX scores per individual per condition. The average NASA-TLX workload score was then plotted with the standard error for each spawn rate and agent delay combination thereby depicting the confidence interval. The results show that for both the 2 and 4 second spawn rates the corresponding longer agent delay times increases the workload. These results are as expected, because at a 2 second spawn rate without automation assistance, the human is more over tasked than for slower spawn rates where the human is able to avoid more collisions and safely land more ships on planets. Workload trends higher with increased agent delay time.



**Figure 3: Spawn, Agent Delay, Workload Analysis**

### Human Draws

A two-factor, repeated measures ANOVA indicated that the effect of the spawn rate had a significant effect on the normalized number of human draws,  $F(1, 9)=61.83$ ,  $MSE=371$ ,  $p>.0001$ ,  $\eta^2 = 0.17$ . The effect size of 0.17 is a small effect. As the effect size is small, it is likely that spawn rate is significant but there is little practical effect of this variable. Furthermore, the normalized human draws measure ANOVA showed that the effect of the automation agent delay time on the normalized number of human draws was significant  $F(1, 9)=62.12$ ,  $MSE=888$ ,  $\eta^2 = 0.41$ ,  $p< 0.0001$ . The effect size as measured with  $\eta^2$  indicates an effect size of medium effect.



**Figure 4: Spawn Rate, Agent Delay, and Human Draws Analysis**

The results show that for spawn rate, as agent delay time increases, the number of normalized human draws increases as depicted in Figure 4. Therefore, the third hypothesis made at the outset of this experiment is supported. The ANOVA showed that the sum of all normalized human draws is significant with respect to the interaction with the agent delay time. The reason for this result could be that as the agent delay time increases, the human is compensating for the ships on screen by making more trajectory draws that were initialized by the automation agent under less demanding conditions, when the probability of collision was lower. Finally, the two-factor repeated measures ANOVA indicates that the interaction between spawn rate and agent delay time has a statistically significant effect on the number of draws  $F(1, 9)=56.78$ ,  $MSE=125$ ,

$p < 0.0001$ ,  $\eta^2 = 0.05$ ; however the effect size is very small at 5%. This effect could be attributed to the fact that during periods of lower workload (4 second spawn rate) the operator is able to draw a larger proportion of the potential routes. Similarly, during periods of high agent delay time, the operator is able to draw a larger proportion of routes as they respond faster than their machine counterpart within the team.

### **Variable Agent Delay with Dynamic Spawn Rate**

The fourth hypothesis surmised that as the task load is decreased, the rate of automation assistance should also be decreased as well allowing the human to stay engaged thus increasing the performance of the human-machine team. For the purposes of Space Navigator, this meant that as the spawn rate was decreased, the agent delay time should also be decreased so that the human-machine team can be optimal. As mentioned previously, the ANOVA indicates that the interaction between spawn rate and agent delay time did not have a statistically significant effect on the normalized score  $F(1, 9) = 1.02$ ,  $MSE = 79.50$ ,  $\eta^2 = 0.0058$ ,  $p < 0.339$ . As the effect size for the interaction of agent delay time with spawn rate is small, these effects are likely not of practical significance. Furthermore, the two-factor repeated measures ANOVA indicates that the interaction between spawn rate and agent delay time did not have a statistically significant effect on the workload  $F(1, 9) = 0.208$ ,  $MSE = 48.15$ ,  $\eta^2 = 0.0027$ ,  $p < 0.339$ .

### **Desirability of Control and Score**

The fifth hypothesis stated that the participant's desirability of control would be inversely related to the normalized score. It was proposed that the higher the desirability of control score, the lower the normalized score would be because the user would be less

prone to relinquishing some control to the automation thereby foregoing the benefits of effective human-machine teaming. Essentially, the operator would compete with the automation by not placing trust in it as a team member and incur a workload burden.

For the desirability of control the variability of the responses to each question was very low. The standard deviation of each response is shown in

Table 3. Each response was very close to 1 point of standard deviation on the scale.

These results show little variability for the responses on the desirability of control scale.

When the total response is calculated, the standard deviation is approximately 9.5 with a mean response of 102, providing a standard deviation of less than 10% of the mean.

Therefore, a lack of variability for the desirability of control questions may indicate that this is not the most useful attribute as an independent variable for this particular group of participants. That is the particular group of participants who participate in this experiment all produced similar desirability of control scores.

**Table 3: Desirability of Control Measures**

Question #	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10	Standard Deviation
1	3	5	4	5	4	6	6	7	5	7	1.317
2	2	3	6	4	5	6	7	1	5	5	1.897
3	2	4	4	5	3	4	3	5	3	6	1.197
4	6	6	6	5	5	5	6	6	3	5	0.949
5	5	6	6	5	4	5	6	6	3	7	1.160
6	5	7	5	4	5	6	2	7	5	3	1.595
7	5	6	6	5	5	6	6	6	6	5	0.516
8	5	6	6	5	6	7	7	7	6	7	0.789
9	3	6	5	5	6	7	7	7	6	7	1.287
10	4	6	6	5	4	4	6	5	2	6	1.317
11	5	4	6	5	6	5	5	4	6	5	0.738
12	5	2	5	5	5	4	6	3	2	3	1.414
13	6	6	5	5	5	6	6	6	5	7	0.675
14	6	6	5	5	5	6	6	7	6	6	0.632
15	4	2	5	5	5	5	6	6	3	4	1.269
16	6	6	5	5	3	4	7	7	6	5	1.265
17	2	6	6	5	5	5	6	7	6	6	1.350
18	3	6	3	5	4	5	6	4	6	5	1.160
19	5	6	4	5	4	3	6	7	5	6	1.197
20	5	4	6	6	3	4	6	6	5	6	1.101
	87	103	104	99	92	103	116	114	94	111	9.569

Linear regression testing showed that there was not a significant correlation between the normalized score and the desirability of control scale results. The results indicate that only 1% of the variance in the normalized score is predicted by the sum of the desirability of control ( $R^2 = 0.01$ ). Therefore, the fifth hypothesis made at the outset of this experiment is rejected.

## **Human Draws and Score**

The sixth and final hypothesis stated that the normalized number of human draws would be directly correlated to the normalized score. Linear regression testing showed that there was not a significant correlation between the normalized score and the number of human draws results;  $R^2 = 0.068$  which means that less than 7% of the variability in score is predicted by the number of human draws. This finding is important because it shows that the number of human draws does not correlate to final score. Different operators responded uniquely to the workload conditions with some of them drawing much more than others to stay engaged with the environment, but this effort had no statistical effect on score. Because of this result, the sixth and final hypothesis made at the outset of this experiment is rejected.

## **Discussion**

Goodman's research showed that score and workload varied as agent delay time changes. While Goodman's experimental results were highly correlated with his model results, and some of these results were confirmed by the current research, there were a few conclusions that Goodman reached that were not supported by research results.

The current research supported Goodman's findings that as agent delay time increases, perceived workload increases, the number of draws completed by the human increases and the actions the human permits the agent to take decreases. Therefore, one can expect that in any system with a well designed automated agent, where the agent's performance is comparable to human performance, performance will also decrease as the rate of automation assistance is decreased.

An interesting disparity appeared between Goodman's results and the results reported in this research. For Goodman's research, the number of human redraws also decreased with increases in agent delay time for the simulation and the HITL experiment. This research supported the hypothesis that the number of normalized human redraws would increase as the agent delay time increased. The ANOVA for the agent delay time, showed that the number of human redraws increased as the agent delay time increased. One possible explanation for the results could be that the current group of users were more likely to relinquish control to the automation which would account for the lower number of redraws during a low agent delay time.

The overall trend in score showed that team performance increased slightly, but not significantly, as a function of decreasing agent delay time. One obvious cause of this research's results in this regard could be the sample size that was used. Similar research conducted by Bindewald and Goodman used as few as 20 and as many as 36 participants. (Bindewald, Peterson, & Miller, 2014) (Goodman, Miller, & Rusnock, 2015) (Goodman, Miller, & Rusnock, 2016). The experiment within this research was conducted with 10 participants. Another more interesting explanation is taken from a result that Goodman noted. One of his conclusions was that by adding automation that drew trajectories for the human, many humans chose to change the underlying task from one of "trajectory drawing" or "ship routing" to a completely different task of "trajectory correction" or "collision avoidance". Goodman found that adaptive automation agent delay settings were crucial in determining the tasks performed by the participants. Automations that triggered at later times trended toward maintaining start-to-finish user engagement, which was similar to the no automation condition (Goodman, Miller, & Rusnock, 2016).

Applying this concept to the results of this research could explain that operators were constantly shifting the task from trajectory drawing, to more frequent collision avoidance induced by the automation drawing more routes. Finally, the most likely result of this relatively insignificant change in performance in the presence of “high” automation rate was that the setting for high automation was incorrect. At the beginning of this research, the assumption was made that the operators would be performing very close to the optimal arousal and performance with help of the high rate of automation assistance. The result that differed from Goodman’s research was that the rate of automation assistance did not have a significant effect on the performance of the human-machine team. On average, the performance of the human-machine team was nearly the same for a high and low rate of automation assistance. These results, show that the high rate of automation assistance may not have been calibrated correctly for the pool of operators that were used in this experiment. The operators of this experiment were never at the optimal arousal and performance peak but were at some place trending toward low arousal. Therefore, they did not increase their redraw rate for shorter agent delay times and the overall team score did not increase for these conditions. Perhaps this lower state of arousal may have resulted from the low spawn rate conditions where the need for user engagement was reduced.

The last hypothesis of this experiment asserted that the operator’s behavior on the environment would be positively correlated to the performance of the human-machine team. Essentially, that the more draws that a human had on the Space Navigator environment, the higher the score would be. The assumption was that the increased number of human draws would result from the human intervention which was more

proficient at the Space Navigator task than the automation assistance as designed for this experiment. Ideally, an increased number of human draws would result in higher performance, however the data collected from Space Navigator showed that this was not supported.

## V. Conclusions and Recommendations

### Investigative Questions Answered

This research set out to investigate six hypotheses regarding automation agent delay time, and human interaction within a game environment. The first hypothesis was that the rate of automation assistance, represented by agent delay time within the game, will have a significant impact on the performance of the human-machine team and the relationship will have a negative correlation. Surprisingly, the analysis on the data showed this hypothesis to be incorrect. Agent delay time did not have a significant effect on the normalized score. The second hypothesis asserted that the workload, as perceived by the participant, would be statistically significant due to the agent delay time. The analysis firmly supported this prediction. In light of the results confirming that agent delay time did not have a significant effect on the normalized score, the decrease in agent delay time made the participants feel less taxed throughout the trials.

The third hypothesis of this research proposed that human engagement, represented by the sum of all human draws in the environment, will be higher during a low rate of automation assistance (i.e., short agent delay times). The analysis showed that for a given spawn rate, as agent delay time increases, the number of normalized human draws increases. The sum of all human draws was significant in the presence of automation and therefore the third hypothesis was confirmed. This third hypothesis was further confirmed by the fact that as the agent delay time increases, the number of human redraws also increased.

The fourth hypothesis stated that as the task load is decreased the rate of automation assistance should also be reduced to maintain similar arousal. It was believed this might be demonstrated through higher performance for increased agent delay times as the spawn rate decreased from one ship every 2s to one ship every 4s. However, interaction effects did not exist for score or workload. The only interaction effect present was for human draws where the interaction illustrated a larger change in the proportion of human draws for the 4s spawn rate than the 2s spawn rate.

The fifth hypothesis proposed that the participants' desirability of control scale measure would be inversely related to the normalized score. The rationale for this was that the user would be less prone to relinquishing control to the automation thereby reaping the benefits of human-machine teaming. The analysis of this research disproved this hypothesis. The last question considered by this research was: "Did the normalized number of human draws positively correlate to the normalized score?" The rationale behind this positive correlation was that with more human draws, the participant will make more decisions that will positively affect the game score which would include more paths drawn to avoid collisions, pick up bonuses, and land ships safely on their respective planets. The analysis of the data showed that this hypothesis was not supported and that the number of human draws had a very small effect on the normalized score; therefore the hypothesis was not supported.

### **Conclusions of Research**

As expected, the results showed large effects of spawn rate on normalized score as well as workload, but in real world systems this increase in task load is not as easily

controlled, and therefore it is not as easy to determine when to engage the automation. Furthermore, this research explored desirability of control and human draw numbers, which were determined to have little to no effect on the normalized score. These results are useful because they provide adaptive automation designers with additional data points in how to approach a new design. While significant performance differences between the groups of individuals who adapt to the agent's timing and those who do not was not fully tested, the general trends in scores indicate that team performance increased as a function of decreasing delay time. Follow on experiments could explore the disparity between those users who adapted to the agent's timing and those who did not and seek to incorporate these results into the model. While measures such as a low desirability of control score or high number of normalized human draws did not result in a statistically significant higher normalized score, perhaps there is interaction with workload. Future analysis on this same data set could explore if there is a correlation between a higher desirability of control score and an increase on workload.

### **Recommendations for Future Research**

One of the surprising results of this research was that the rate of automation assistance did not have a significant effect on performance, but it did have a significant effect on workload. Future experiments could focus on testing multiple time increments shorter and longer than 5 minutes to explore the effects of short duration and prolonged high workload and rate of automation assistance effects on performance. The idea behind this is that the five-minute time trials are just one timeframe and may not be the most accurate depiction to observe a significant difference in the normalized score. Future

research should perform multiple regression analysis including spawn rate, agent delay time, and human draws as independent variables in the regression. This analysis could show to what extent these factors are responsible for performance in the system.

While the adaptive automation model was able to capture workload and performance to a reasonable degree of certainty, future research would also model operator engagement. The introduction of automation inherently causes a change in the task load for the operator. Therefore, future adaptive automation models should incorporate modelling of how operators adjust their engagement with the system as a function of the automation agent delay time.

### **Significance of Research**

The results of analysis like this, could show that human performance improvements as well as task load will need to be accounted for in the design of adaptive automation systems. The task load and the human response rate would contribute to a better designed adaptive automation. For instance, this experiment showed that automation assistance was not as helpful when set to the predetermined value of 2.6 seconds as for previous studies while the task load of 2, 3, and 4 seconds were used as in previous work. Perhaps this discrepancy was due to a lower rate of overall arousal for the present participants. Such a state might indicate that the participants in this study were higher skilled than participants in earlier studies or that the lower spawn rate led to this overall decrease in arousal, resulting in less motivated participants. As designers make decisions about adaptive automation they will need to consider both the human response rate, an indicator of engagement, and where it lies on the Yerkes Dodson curve.

Once user engagement is accounted for, designers can proceed to adjust the automation for optimal human-machine teaming that provides adequate engagement and performance.

## Appendix A - Game Play Pre-Questionnaire

FOR OFFICIAL USE ONLY

### Game-Play Data Collection Study Pre-Questionnaire

1. Participant Number (Assigned By Researcher):

2. Age:

3. Handedness:

4. Gender:

5. Have you previously participated in a *Space Navigator* experiment?      YES      NO

If YES, please explain which one:

6. How much experience do you have with the following:

a. In a given year, on how many days do you interact with the following types of devices?

	Never	<1 per month	1-3 per month	1-2 per week	3-6 per week	Daily
Laptop computer	0	1	2	3	4	5
Tablet computer	0	1	2	3	4	5
Smart phone	0	1	2	3	4	5
Desktop computer	0	1	2	3	4	5
Gaming consoles	0	1	2	3	4	5

a. In a given year, on how many days do you play **video games** of the following type?

	Never	<1 per month	1-3 per month	1-2 per week	3-6 per week	Daily
Simulation (SimCity)	0	1	2	3	4	5
Role-Playing (WoW)	0	1	2	3	4	5
Action (Mario, Donky Kong)	0	1	2	3	4	5
First Person Shooter (Halo)	0	1	2	3	4	5
Strategy (Civilization)	0	1	2	3	4	5
Puzzle (Tetris, Candy Crush)	0	1	2	3	4	5
Casual (Angry Birds)	0	1	2	3	4	5
Music (Guitar Hero)	0	1	2	3	4	5
Sports (Madden Football)	0	1	2	3	4	5
Board (Monopoly)	0	1	2	3	4	5
Card (Poker, Pinochle)	0	1	2	3	4	5
<b>All Video Games (TOTAL)</b>	0	1	2	3	4	5

## Appendix B – Desirability of Control

7. Below you will find a series of statements. Please read each statement carefully and respond to it by expressing the extent to which you believe the statement applies to you. For all items, a response from 1 to 7 is required. Circle the number that best reflects your belief using the following scale:

1 = The statement does not apply to me at all

2 = The statement usually does not apply to me

3 = Most often, the statement does not apply

4 = I am unsure about whether or not the statement applies to me, or it applies to me about half the time

5 = The statement applies more often than not

6 = The statement usually applies to me

7 = The statement always applies to me

1. I prefer a job where I have a lot of control over what I do and when I do it.

1 2 3 4 5 6 7

2. I enjoy political participation because I want to have as much of a say in running government as possible.

1 2 3 4 5 6 7

3. I try to avoid situations where someone else tells me what to do.

1 2 3 4 5 6 7

4. I would prefer to be a leader than a follower.

1 2 3 4 5 6 7

5. I enjoy being able to influence the actions of others.

1 2 3 4 5 6 7

6. I am careful to check everything on an automobile before I leave for a long trip.

1 2 3 4 5 6 7

7. Others usually know what is best for me.

1 2 3 4 5 6 7

8. I enjoy making my own decisions.

1 2 3 4 5 6 7

9. I enjoy having control over my own destiny.

1 2 3 4 5 6 7

10. I would rather someone else take over the leadership role when I'm involved in a group project.

1 2 3 4 5 6 7

11. I consider myself to be generally more capable of handling situations than others are.

1 2 3 4 5 6 7

12. I'd rather run my own business and make my own mistakes than listen to someone else's orders.

1 2 3 4 5 6 7

13. I like to get a good idea of what a job is all about before I begin.

1 2 3 4 5 6 7

14. When I see a problem, I prefer to do something about it rather than sit by and let it continue.

1 2 3 4 5 6 7

15. When it comes to orders, I would rather give them than receive them.

1 2 3 4 5 6 7

16. I wish I could push many of life's daily decisions off on someone else.

1 2 3 4 5 6 7

17. When driving, I try to avoid putting myself in a situation where I could be hurt by another person's mistake.

1 2 3 4 5 6 7

18. I prefer to avoid situations where someone else has to tell me what it is I should be doing.

1 2 3 4 5 6 7

19. There are many situations in which I would prefer only one choice rather than having to make a decision.

1 2 3 4 5 6 7

20. I like to wait and see if someone else is going to solve a problem so that I don't have to be bothered with it.

1 2 3 4 5 6

## Appendix C – Game Play Post-Questionnaire

### Game-Play Data Collection Study Post-Questionnaire

1. Participant Number (Previously Assigned By Researcher):
2. In your opinion, how well did you play *Space Navigator*?
3. Did the adaptive automation aids help or hinder your ability to play *Space Navigator* well? Please explain.
4. How did the way you played *Space Navigator* change as you progressed through the experiment?
5. What strategies did you adopt over time that helped you play *Space Navigator* more successfully?
6. What outside factors may have influenced your ability to perform well?
8. What changes would you make to the automated aids in order to perform better in future games?
10. Which version was easiest? Hardest?
11. What level of trust did you have in the automated aids?
12. What affected your perception of trust in the system positively or negatively? Why?
13. Which condition do you think would permit you to perform the best over a longer period of time?

## Appendix D: RStudio Two Way Repeated Measures Code

```
#-----Draw/Redraw AOV-----
aov_redraws<-read.csv("/Users/markharris/Desktop/R_2_Way_Repeated_Anova_2x2_mean2.csv", header = T) #Read in the csv file
aov2_redraws <- within(aov_redraws, {
  fSubject <- factor(Subject)
  fAgentDelay <- factor(Trigger)
  fTrigger <- factor(Trigger)
  fSpawn <- factor(Spawn)
  fNormalization <- factor(Normalization)
  fAgentDraws <- factor(NormInitialAutoDraw)
  fInitialHumanDraws <- factor(NormInitialHumanDraw)
  fNormAllHumanRedraw <- factor(NormAllHumanDraws)
  fNormHumanRedraw <- factor(NormHumanRedraw)
  fNormAutoRoutesRedrawn <- factor(NormAutoRoutesRedrawn)
  fDesirabilityofControlScale <- factor(DesirabilityofControlScale)
})
#factor the columns for the ANOVA

Two_way_aov_workload <- aov(Workload ~ fSpawn * fTrigger + Error(fSubject / (fSpawn * fTrigger)), data = aov2_redraws)
#Perform 2 way repeated anova with error calculation
summary(Two_way_aov_workload) #Summarize the results

Two_way_aov_agent_draws <- aov(InitialAutomationDraws ~ fSpawn * fAgentDelay + Error(fSubject / (fSpawn * fAgentDelay)),
data = aov2_redraws) #Perform 2 way repeated anova with error calculation
summary(Two_way_aov_agent_draws) #Summarize the results

Two_way_aov_human_draws <- aov(NormAllHumandraws ~ fSpawn * fAgentDelay + Error(fSubject / (fSpawn * fAgentDelay)),
data = aov2_redraws) #Perform 2 way repeated anova with error calculation
summary(Two_way_aov_human_draws) #Summarize the results

Two_way_aov_initial_human_draws <- aov(InitialHumanDraws ~ fSpawn * fAgentDelay + Error(fSubject / (fSpawn *
fAgentDelay)), data = aov2_redraws) #Perform 2 way repeated anova with error calculation
summary(Two_way_aov_initial_human_draws) #Summarize the results

Two_way_aov_Norm_Human_Redraw <- aov(NormHumanRedraw ~ fSpawn * fAgentDelay + Error(fSubject / (fSpawn *
fAgentDelay)), data = aov2_redraws) #Perform 2 way repeated anova with error calculation
summary(Two_way_aov_Norm_Human_Redraw) #Summarize the results

Two_way_aov_auto_routes_redraws <- aov(NormAutoRoutesRedrawnByHuman ~ fSpawn * fAgentDelay + Error(fSubject /
(fSpawn * fAgentDelay)), data = aov2_redraws) #Perform 2 way repeated anova with error calculation
summary(Two_way_aov_auto_routes_redraws) #Summarize the results

Two_way_aov_DesirabilityControl <- aov(DesirabilityofControlScale ~ fSpawn * fAgentDelay + Error(fSubject / (fSpawn *
fAgentDelay)), data = aov2_redraws) #Perform 2 way repeated anova with error calculation
summary(Two_way_aov_DesirabilityControl) #Summarize the results
```

## Bibliography

- Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-Centered Design. In W. Bainbridge, *Berkshire Encyclopedia of Human-Computer Interaction* (pp. 445-456). Thousand Oaks: Sage Publications.
- Billings, C. (1991). *Human-centered aircraft automation: A concept and guidelines*. NASA-Ames Research Center. Moffet Field, CA: NASA Tech. Memo. No. 103885.
- Billings, C. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ: Lawrence Erlbaum Assoc.
- Bindewald, J. M. (2015, September 17). *Adaptive Automation Design and Implementation*. Retrieved from Defense Technical Information Center: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA623873>
- Bindewald, J. M., Miller, M., & Peterson, G. (2014). A function-to-task process model for adaptive automation system design. *International Journal of Human-Computer Studies*, 13.
- Bindewald, J. M., Peterson, G. L., & Miller, M. E. (2014, April 30). Creating Effective Automations that Maintain Explicit User Engagement. Wright-Patterson AFB, OH, USA.
- Dzindolet, M., Peterson, S., Pomranky, R., Pierce, L., & Beck, H. (2003). The role of trust in automation reliance. *International Journal of Human Computer Studies*, 58(6), 697-718.
- Endsley, M. (1987). The application of human factors to the development of expert systems for advanced cockpits. *Proceedings of the Human Factors Society 31st Annual Meeting* (pp. 1388-1392). Santa Monica: Human Factors and Ergonomics Society.
- Endsley, M. R. (1988). Situational awareness global assessment technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference*, (pp. 789-795). Dayton.
- Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman, & M. Mouloua, *Automation and human performance performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum Associates.

- Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 381-394.
- Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*(42), 462-469.
- Goodman, T., Miller, M., & Rusnock, C. (2015). Incorporating Automation: Using Modeling and Simulation to Enable Task Re-allocation. *Winter Simulation Conference* (p. 12). Huntington Beach, CA: IEEE.
- Goodman, T., Miller, M., & Rusnock, C. (2016). Timing Within Human-Agent Interaction and its Effects on Team Performance and Human Behavior. *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (p. 7). San Diego: IEEE.
- Goodman, T., Miller, M., Rusnock, C., & Bindewald, J. (2017). Effects of Agent Timing on the Human-Agent Team. *Cognitive Systems Research*, 17-33.
- Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1997). Operator versus computer control of adaptive automation. *Proceedings of the 7th International Symposium on Aviation Psychology* (pp. 161-166). Columbus: Department of Aviation, The Ohio State University.
- Itoh, M. (2011). A model of trust in automation: Why humans over-trust. *IN: SICE Annual Conference (SICE)* (pp. 198-201). Tokyo: Proceedings of IEEE.
- Kaber, D. B., & Endsley, M. (2003). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 1-40.
- Kaber, D., Riley, J., Tan, K.-W., & Endsley, M. (2001). On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics*, 37-57.
- Kessel, C., & Wickens, C. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors*, 24, 49-60.
- Maria, A. (1997). Introduction to Modeling and Simulation. *Proceedings of the 1997 Winter Simulation Conference* (pp. 1 - 13). Binghamton, NY: State University of New York at Binghamton.

- Norman, D., & Draper, S. (1986). *User-Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Parasuraman, R. (1993). Effects of adaptive function allocation on human performance. In D. G. (eds), *Human factors and advanced aviation technologies* (pp. 147-158). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.
- Parasuraman, R. M. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 665-679.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse, disuse and abuse. *Human Factors*, 230-253.
- Pope, A. T., Comstock, R. J., Bartolome, D. S., Bogart, E. H., & Burdette, D. W. (1994). Biocybernetic system validates index of operator engagement in automated task. In M. Mouloua, & R. Parasuraman, *Human Performance in automated systems; Current research and trends* (pp. 300-306). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Sheridan, T. (1997). Supervisory Control. In G. Salvendy, *Handbook of human factors* (pp. 1295-1327). New York, NY: John Wiley and Sons.
- Sheridan, T., & Parasuraman, R. (2006). Human-automation interaction. In R. Nickerson, *Review of human factors and ergonomics* (pp. 89-129). Santa Monica, CA: Human Factors and Ergonomics Society.
- Steinhauser, N. P. (2009). Design Principles for Adaptive Automation and Aiding. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 6-10.
- Wickens, C., & Wickens, C. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. *Human Factors*, 49-60.
- Wiener, E. L. (1988). Cockpit Automation, in E.L. Wiener and D.C. Nagel (eds). *Human Factors in Aviation*, 433-459.
- Young, L. (1969). On adaptive manual control. *Ergonomics*, 12, 635-657.

<b>REPORT DOCUMENTATION PAGE</b>			Form Approved OMB No. 074-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p><b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b></p>				
<b>1. REPORT DATE (DD-MM-YYYY)</b> 06-15-2017		<b>2. REPORT TYPE</b> Master's Thesis		<b>3. DATES COVERED (From - To)</b> March 2012 - May 2017
<b>TITLE AND SUBTITLE</b>  Variable Timing Effects on Performance and Behavior within Human-Machine Teams			<b>5a. CONTRACT NUMBER</b>	
			<b>5b. GRANT NUMBER</b>	
			<b>5c. PROGRAM ELEMENT NUMBER</b>	
<b>6. AUTHOR(S)</b>  Harris, Mark A., Captain, USAF			<b>5d. PROJECT NUMBER</b>	
			<b>5e. TASK NUMBER</b>	
			<b>5f. WORK UNIT NUMBER</b>	
<b>7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)</b> Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-7765			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>  AFIT-ENV-MS-17-J-056	
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> US Air Force Office of Scientific Research, Mathematical and Computational Cognition Program Attn: James H. Lawton, PhD 4075 Wilson Blvd., Suite 350 Arlington, VA 22203 <a href="tel:7036965999">(703) 696-5999</a> (DSN: 426-5999) <a href="mailto:James.Lawton.1@us.af.mil">James.Lawton.1@us.af.mil</a>			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>  AFOSR/RTA2	
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>	
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> <b>DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.</b>				
<b>13. SUPPLEMENTARY NOTES</b> This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.				
<b>14. ABSTRACT</b>  When well designed human-agent teaming is used within a system, it provides a tremendous benefit than neither the human or automation can accomplish alone. This research sought to explore various parameters that would help in the effective design of robust adaptive automation. Agent timing and its interaction with spawn rate were explored from a performance perspective and to understand the number of times a human engages with the environment. Surprisingly, agent timing or certain user actions within the environment did not significantly affect performance but significant changes in workload were observed. Adaptive automation design seeking to maximize human-agent team performance should have a thorough understanding of the human experimentation results needed to explore the effect of an artificial agent's timing on the performance of a human-agent team within a highly dynamic task environment. The research contained within explores those human experimentation results.				
<b>15. SUBJECT TERMS</b> Automation, Adaptive Automation, Human Performance				
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>  UU	<b>18. NUMBER OF PAGES</b>  63
<b>a. REPORT</b>  U	<b>b. ABSTRACT</b>  U	<b>c. THIS PAGE</b>  U		
			<b>19b. TELEPHONE NUMBER (Include area code)</b> (937) 255-6565, ext 4651 ( <a href="mailto:michael.miller@afit.edu">michael.miller@afit.edu</a> )	