3-21-2019

Assessing Artificial Agent Response Time Effects on Human-Agent Teams in Variable Inter-Arrival Time Environments

David J. Canzonetta

ASSESSING ARTIFICIAL AGENT RESPONSE TIME EFFECTS ON HUMAN-AGENT TEAMS IN VARIABLE INTER-ARRIVAL TIME ENVIRONMENTS

THESIS

David J. Canzonetta, First Lieutenant, USAF

AFIT-ENV-MS-19-M-166

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-MS-19-M-166

ASSESSING ARTIFICIAL AGENT RESPONSE TIME EFFECTS ON HUMAN-AGENT TEAMS IN VARIABLE INTER-ARRIVAL TIME ENVIRONMENTS

THESIS

Presented to the Faculty

Department of Systems Engineering and 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

David J. Canzonetta, BS

First Lieutenant, USAF

March 2019

AFIT-ENV-MS-19-M-166

ASSESSING ARTIFICIAL AGENT RESPONSE TIME EFFECTS ON HUMAN-AGENT TEAMS IN VARIABLE INTER-ARRIVAL TIME ENVIRONMENTS

David J. Canzonetta, BS
First Lieutenant, USAF

Committee Membership:

Michael E. Miller, PhD
Chair

John M. McGuirl, PhD
Member

Gilbert L. Peterson, PhD
Member

Acknowledgments

I would like to thank my research advisor, Dr. Miller, for his unwavering support during the formulation of this thesis. The timely and honest quality feedback (and there was a lot that I requested) made this work drastically more gratifying. I would also like to thank Dr. McGuirl for the big-picture navigation and guidance of this research at points where I thought it had hit a wall.

David J. Canzonetta

Table of Contents

Page

List of Figures

Page

List of Tables

AFIT-ENV-MS-19-M-166

Abstract

Autonomous systems have gained an expanded presence within the Department of Defense (DoD). Furthermore, the DoD has clearly stated autonomous systems must extend the capabilities of their human operators. Thus, the exploration of strategies for effective pairing of humans and automation supports this vision. Previous research demonstrated that the time at which an automated agent assumes a task for its human teammate, or agent response time (ART), affects human-agent team performance, human engagement, and human workload. However, in this research environment, the time between subsequent tasks appearing to the human-agent team, or inter-arrival time (IAT), remained constant. Variable IAT environments more accurately reflect real-world operational environments. Previous research also maintained ART at a fixed level. Additionally, the effect of human understanding of automated teammate actions on human-agent team performance remains unknown.

This thesis attempts to analyze the effect of an agent with adaptive ART that varies based on current IAT on human-agent team performance, human engagement, and human workload. Additionally, it seeks to determine the implication of agent predictability to the human. This thesis explores these issues in three phases. First, a method and development of a variable ART function for use in future phases is presented. Second, a study of a variable ART teammate against a fixed ART teammate highlights the significance of providing detailed agent instruction to the human. Third, analysis of instruction and type of agent teammate across an entire input IAT function and at different IAT levels is conducted. This work establishes key factors for adaptive ART function implementation.

Based on specific IAT changes, the current research demonstrates that adaptive ART can boost human-agent team performance and manipulate human engagement. Furthermore, predictability of agent action in variable IAT environments is a desired system attribute.

ASSESSING ARTIFICIAL AGENT RESPONSE TIME EFFECTS ON HUMAN-AGENT TEAMS IN VARIABLE INTER-ARRIVAL TIME ENVIRONMENTS

I. Introduction

General Issue

Autonomous systems promise the ability to boost the pace of operations, decrease frivolous labor costs, reduce operational launch time, increase operational reliability, and remove the human operator from imminent danger (Department of Defense: Defense Science Board, 2012; Endsley, 2015). As illustrated in the fiscal year 2018 (FY18) Department of Defense (DoD) budget, the benefits of autonomous systems impact the domains of ground, maritime, and air (Gettinger, 2017).

Autonomous systems have gained an expanded presence within the DoD. For example, the FY18 Unmanned Aerial Vehicle (UAV) development budget in the DoD totaled $6.97 billion and represented a five-year high (Gettinger, 2017). The DoD spent a total of $34.6 billion developing unmanned systems from fiscal year 2013 to FY18 (Gettinger, 2017). Furthermore, the Unmanned Systems Integrated Roadmap 2017-2042 demonstrates the DoD intent to increase emphasis on unmanned systems (Fahey & Miller, 2017). Additionally, the United States Air Force (USAF) has explored the concept of creating a UAV that can serve as a wingman for a fighter aircraft (Kearns, 2015). DoD autonomous systems can also take the form of ground-based systems. For instance, the Army's BigDog project existed to create a robotic mule to carry the packs of ground soldiers (Raibert, Blankespoor, Nelson, & Playter, 2008). Although high noise factors ceased its ten-year development, the BigDog project still provides one example of the DoD desire to place more emphasis on autonomous systems.

Autonomous systems can provide these benefits using agents. An agent is “a kind of physical object that can act in the environment, perceive events, and reason” (Sterling & Taveter, 2009). Furthermore, an autonomous agent is “a kind of agent that creates and pursues its own agenda as opposed to functioning under the control of another agent” (Sterling & Taveter, 2009). Within the scope of DoD missions, concerns of independent autonomy employing lethal methods may cause a shift from this definition. Autonomous components of DoD systems represent an example of “controlled autonomy,” meaning they operate on someone’s behalf (Sterling & Taveter, 2009).

DoD autonomous systems seek to enhance human ability to successfully complete missions (Department of Defense: Defense Science Board, 2012). For this enhancement to occur, the human must maintain a strong presence within autonomous systems. A favorable working relationship between the human and automated agents enables a strong human presence and a high-level of performance. Human-agent team performance is expressed as the level of success a human and agent attain while striving to accomplish the system goal. Additionally, human engagement is the amount of human involvement within a human-agent team. It is desired to keep the human engaged in the present task because, in too much of a supervisory role, the human requires more effort to maintain acceptable levels of vigilance and alertness (Goodman, Miller, Rusnock, & Bindewald, 2017; Parasuraman, 2008). Furthermore, proper control of human workload is often desired. Human workload is the “the impact of the task demand placed upon the operator’s mental or physical resources" (Bindewald, Miller, & Peterson, 2014). Ideally, workload remains at a level that engages but does not overwhelm the human operator. These examples demonstrate that human-agent team performance, human engagement, and human workload are factors influenced by human-agent interaction.

Previous research at the Air Force Institute of Technology (AFIT) investigated the effect of the time at which automation assumes a task for the human on the response variables of human-agent team performance, human engagement, and human workload (Goodman et al., 2017). This research concluded that autonomous agent timing significantly affects these response variables (Goodman et al., 2017). Autonomous agent timing, also referred to as "agent response time (ART)," is the amount of time elapsed before an agent becomes involved in a task. Further research determined the impact of different task inter-arrival times paired with certain ARTs on human-agent team performance, human engagement, and human workload (Schneider, Bragg, Henderson, & Miller, 2018). Inter-arrival time (IAT) is the amount of time between new tasks appearing to the human-agent team. This research recommended a potential function for calculating ART based on IAT (Schneider et al., 2018). Both studies used a constant IAT environment and opened questions regarding the proper ART in variable IAT environments.

The studies conducted by Goodman et al. and Schneider et al. used a tool developed at AFIT titled "Space Navigator" to conduct their experiments (Goodman et al., 2017; Schneider et al., 2018). Space Navigator is an air-traffic management game (Bindewald et al., 2014). The game places the human user on a team with an automated agent. The team attempts to achieve the highest game score possible. The human-agent team obtains their score by navigating ships that spawn onto the screen to their correct destination. The automated agent can create an initial path from a ship to a planet. The provided agent path includes the potential for collisions with other ships. Space Navigator enables researchers to measure the output variables of human-agent team performance, human engagement, and human workload.

Problem Statement

The impact of autonomous agent timing in variable IAT environments remains unknown. While previous research demonstrated that autonomous agent timing does have an effect on performance, it only explored this effect in a constant IAT environment (Goodman et al., 2017). Variable IAT environments better represent real-world operating settings. Additionally, the actions of the agent in the previous studies remained the same regardless of changes happening within the context of the environment (Goodman et al., 2017). Adaptive autonomous agent timing may serve as an alternative to traditional function allocation methods to maximize human-agent team performance while maintaining appropriate human engagement and workload balance.

Prior research determined that humans desire predictability of their automated teammate actions (Bindewald, Miller, & Peterson, 2019). However, the impact of providing explicit cues to permit the human to predict ART remains unknown. Providing instruction to the human could impact the ability of the human and agent to work together and thereby maximize team performance.

Research Questions

This thesis attempts to answer the research questions listed below:

- How do we determine a method for effective timing of an artificial agent within a variable IAT environment?
- How do IAT and ART relate to human-agent team performance, human engagement, and human workload?
- How does a variable ART teammate, as compared to a fixed ART teammate, affect human-agent team performance, human engagement, and human workload?
- How does IAT level affect human-agent team performance and human engagement?
- How does explanation of agent functionality to a human affect human-agent team performance, human engagement, and human workload?
- How do humans view a variable ART teammate compared to a fixed ART teammate in terms of predictability?

Scope

This thesis expanded upon research conducted by 2d Lt Tyler Goodman by studying the effect of autonomous agent timing in variable IAT environments (Goodman, 2016). Variable IAT environments more accurately mirror real-world operating environments than constant IAT environments.

Research conducted by Schneider and colleagues provided the model for the experiment structure (Schneider et al., 2018). They attempted to use the Space Navigator environment to establish a basic function to calculate ART as a function of IAT (Schneider et al., 2018).

The test environment consisted of a clean desk setup in a quiet and secluded location. The participant had the ability to adjust his or her seat to a desired comfort level. Lighting remained at its normal daily intensity. The participant played Space Navigator on a Microsoft Surface Pro 4 using his or her finger to manipulate objects within the game.

Methodology Overview

The Space Navigator application provided the environment for thesis experiments. This thesis contained two experiments. The first experiment served to identify an equation to calculate ART as a function of IAT. The first experiment employed an environment with constant ART and constant IAT. Suggested values in previous research guided the selection of specific ART and IAT combinations (Schneider et al., 2018). Optimization techniques applied to results of the first experiment identified a calculation of ART as a function of IAT. The second experiment used this equation as an input value.

The second experiment employed a variable IAT environment. The second experiment compared ART calculated as a function of IAT (as identified in the first experiment) to a constant ART. Furthermore, it explored the impact of providing an explanation of agent functionality to the human. Two-factor, mixed-design analysis of variances (ANOVA) investigated potential significance between independent variables (teammate type and instruction of agent functionality) and dependent variables (human-agent team performance, human engagement, and workload) across the entire input IAT function. Three-factor, mixed-design ANOVAs investigated potential significance between independent variables (teammate type, instruction of agent functionality, and IAT level) and dependent variables (human-agent team performance and human engagement) at different IAT levels. Subsequent one-factor ANOVAs investigated interacting effects.

Assumptions/Limitations

One limitation of this research is that the Space Navigator environment represents a more fleeting and intuitive scenario than a typical operating environment. Moreover, the innocuous consequences of a human operator mistake in the Space Navigator environment contrast with the life or death consequences typically seen in a military environment. However, Space Navigator proves useful by delivering a well-regulated and event-driven environment that provides adjustable IAT and tracks human engagement (Goodman et al., 2017). Thus, this research assumes that the Space Navigator environment will provide results applicable to real-world military operating environments.

The potential differences between human participants in this study and the population of military operators represent another research limitation. One could expect a difference in decision-making strategies between experiment participants and military operators. However, the intuitive play style of Space Navigator makes it easy to train participants (Goodman et al., 2017). Thus, this research assumes that experiment participants will provide results applicable to military operators.

The random nature of ship spawn location within the experiment environment of Space Navigator presents another potential research limitation. Initial starting locations of tasks differ across rounds of Space Navigator. Conceivably, some rounds of Space Navigator could present differing amounts of complex tasks to the user than other rounds. However, like flipping a coin many times in succession, the random nature of task generation should balance tasks of differing complexities over the entire course of experiments. Thus, this research assumes that trials with identical input variables within the experiment environment of Space Navigator have the same degree of complexity.

Materials and Equipment

This research required a Microsoft Surface Pro 4 tablet, which housed the Space Navigator application environment. The Department of Electrical and Computer Engineering at AFIT completed the development of Space Navigator (Bindewald, 2015), while Mr. Derek Desentz augmented game updates.

Thesis Outline

This thesis contains three interconnected articles that provide in-depth details and discoveries from this research. Chapter II contains the article titled “Identifying a Possible Function for Artificial Agent Adaptation in Variable Task Rate Environments” (article in press for International Symposium on Aviation Psychology). This chapter overviews an initial experiment and analysis conducted to identify a function for calculating ART as a function of IAT. Chapter III contains the article titled “Adaptive Artificial Agent Response Time Impact on Human-Agent Team Performance” (article in press for the Institute of Industrial and Systems Engineering Annual Conference and Expo 2019). This chapter discusses the results and recommendations from a portion of the experiment that employed the ART function identified in Chapter II. Chapter IV contains the article titled “Impact of Automated Agent Timing in Variable Inter-Arrival Time Environments” (article in revision for Journal Submission). This chapter re-stated the steps for ART function formulation identified in Chapter II and built upon experiments employed in Chapter III through the introduction of explanation of agent functionality to the human. Chapter V serves to conclude the thesis by answering the research questions, providing direction for future research, and reiterating the significance of current research.

II. Identifying a Possible Function for Artificial Agent Adaptation in Variable Task Rate Environments

Description

Chapter II consists of a conference paper that identifies a specific method for calculating agent response time as a function of inter-arrival time. The function identified in this chapter is utilized as an input to experiments conducted in subsequent chapters. Chapter II provides answers to the first and second research questions listed in Chapter I.

Publication Details

The International Symposium for Aviation Psychology accepted the article provided in this chapter for publication in its conference proceedings. The conference will take place May 7-10, 2019 in Dayton, Ohio.

IDENTIFYING A POSSIBLE FUNCTION FOR ARTIFICIAL AGENT ADAPTATION IN VARIABLE TASK RATE ENVIRONMENTS

David J. Canzonetta and Michael E. Miller
Air Force Institute of Technology
Wright-Patterson, AFB, OH

The current research sought to identify a method to calculate agent response time (ART) as a function of inter-arrival time (IAT), which balances human-agent team performance, human engagement, and human workload. A human-in-the-loop experiment evaluated human-agent team performance, as measured by team score, human engagement, as measured by the number of manually performed tasks, and workload, as measured through a subjective questionnaire, as a function of IAT and ART combination. Results demonstrated that task IAT strongly correlated with performance, engagement, and workload, while ART strongly related to engagement. Optimization was applied to the resulting data to determine ARTs which maximized performance while sustaining desirable levels of human engagement and workload. The optimization produced an ART function for application in future research to judge the effectiveness of adapting ART to boost human-agent team performance.

Humans and artificial agents can be teamed together to complete intricate and vital tasks. Successful task completion relies on the balance of human engagement and workload within these teams. For example, an unengaged human operator experiencing underload can face decreased alertness (Parasuraman, 2008). Dynamic function allocation is a common adaptive automation method for maintaining proper workload balance (Schneider, Bragg, Henderson, & Miller, 2018). However, this type of function allocation can force the human to maintain awareness of their present tasks within the current allocation, effectively increasing mental workload (Kaber, Riley, Tan, & Endsley, 2001).

Previous research conveyed that agent responsiveness within a human-agent team can affect human engagement (Goodman, Miller, Rusnock, & Bindewald, 2017). This discovery suggests that a well-timed agent response could provide an alternative approach to achieving the proper balance between human engagement and human workload in systems employing adaptive automation. For situations where environmentally-imposed inter-arrival time (IAT) heavily influences operator workload, calculation of optimal agent response time (ART) as a function of IAT becomes a possible method for task load sharing. The current study varied IAT and ART, measuring their effects on human-agent team performance, human engagement, and human workload. The data collected from this study produced a function for desired ART as a function of IAT to support future research.

Method

Participants

The experiment involved 14 participants (9 male and 5 female). Two participants were left-handed. Mean participant age was 25.4 and ranged from 20 to 31. One participant had previous Space Navigator experience. All but one participant exhibited normal color vision using the Ishihara Color Deficiency Charts (Ishihara, 2012). The participant with apparently irregular color vision obtained the third highest recorded score, indicating their ability to successfully identify the items in the game. Therefore, the analysis included their data. Participants self-reported spending an average of 48.7 hours per week using a computer or similar machine.

Apparatus and Environment

The experiment used a touch-screen tablet application titled "Space Navigator." Space Navigator closely resembles commercially-available air-traffic-control games. In this game, a human and agent work together as peers to achieve the highest score possible. The object of Space Navigator is to navigate red, blue, yellow, or green ships that spawn onto the screen to planets of their corresponding color, while obtaining randomly-appearing bonuses during their routes. The human-agent team receives 100 points upon successful navigation of ships to their corresponding planet. Ships are removed from the screen when they arrive at their appropriate planet. Additionally, the human-agent team receives 50 points for navigating ship paths through bonuses that appear on the screen. A bonus appears at a random on-screen location once every 10 seconds and remains on-screen until collected by a ship. The team loses 200 points when two ships collide. The human can physically draw a ship path with their finger, but if the human does not draw a path within a specified time window, the artificial agent presents a straight-line path from the ship to its appropriate planet. However, this agent path does not account for any bonuses or the paths of any other ships on screen. The human can draw or redraw a route at any time. The agent cannot overwrite a human-drawn route. Participants played all games on a Microsoft Surface Pro 4 in a quiet and secluded location.

Experimental Design and Procedure

The input variables to this study were agent response time (ART) and inter-arrival time (IAT). ART is the time an agent waits to draw a route for a new ship. IAT is the number of seconds between the times that two subsequent ships appear. Previous research narrowed and tested a range of IAT and ART values from 2s to 4s and 2.6s to 8.6s, respectively (Schneider et al., 2018). This research analyzed how the ratio of ART to IAT, referred to as the Adaptation Coefficient (AC), affects score, engagement, and workload (Schneider et al., 2018).

Decreasing IATs result in more ships appearing within a given time. This has the apparent and desired effect of increasing task load by requiring the human-agent team to provide more routes within a given time interval. Since these ships remain in the environment for a period of time to transit to their destination planet, the density of ships in the environment increases, increasing the probability of collisions, and reducing the number of possible collision free routes within the environment. This effect further increases task load as the human must draw or redraw longer and more complex routes.

Figure 1 displays the IAT and ART points used in this experiment, illustrated by points with markers “x” and “o”, respectively. Past studies narrowed the sampling area to boundaries and points featured in Figure 1 by demonstrating team performance in the experiment environment remained similar for IAT values greater than 3.4s (Goodman et al., 2017; Schneider et al., 2018). The dashed lines that create the top and bottom boundaries represent AC of 2.0 and 0.5, respectively. These AC were chosen because they represent locations of manageable human workload in the Space Navigator environment, as discovered in previous research (Schneider et al., 2018), although human-agent team performance varied within this range. When IAT is significantly less than 2.6s, the human will struggle to keep up with new tasks, thereby experiencing overload. When IAT is significantly greater than 2.6s, the human will experience large breaks between new tasks, thereby experiencing underload. As ART decreases, the human typically draws routes slower than the agent, which could prevent the human from drawing and thereby decrease human engagement. Conversely, as ART increases, the human can draw routes faster than the agent, so one might assume that human engagement increases.

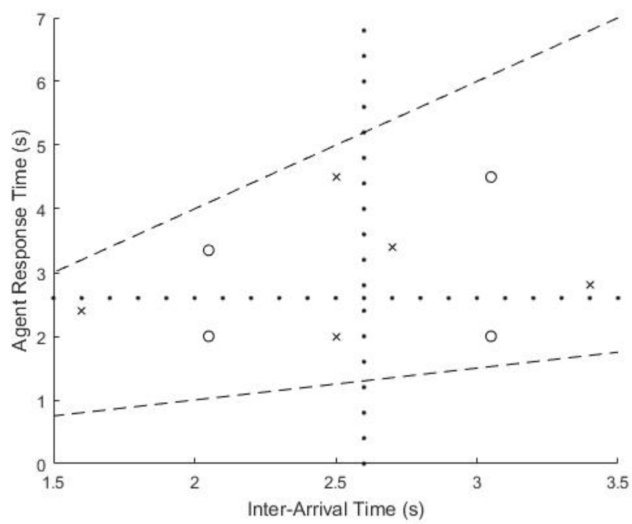

Figure 1. Depiction of Inter-Arrival Time (IAT) and Agent Response Time (ART) points sampled during the current experiment (shown as x’s and o’s). The vertical and horizontal dotted lines indicate the average human draw time of 2.6 s. The sloped dashed lines indicate a range of values useful for human-machine teaming based on previous research. Points marked with an “o” in Figure 1 represent the centroid of each region within the boundaries provided by the dashed and dotted lines. Points marked with an “x” were selected to be near the boundary extremes to provide insight into human performance near these transition regions.

For each experimental session, the research administrator provided a demonstration of Space Navigator to participants from a narrated script. The participants then played three, 2.5 minute practice rounds, each with an agent teammate, to become familiar with the Space Navigator environment. Practice rounds contained slower than average IAT and ART values to give participants time to understand the mechanics of the game and touchscreen response. Participants received no gameplay strategies during training.

The experimental session for each participant contained two blocks. Each block included nine, 1.75 minute trials with a workload questionnaire administered after each trial. Game time remained constant in all trials. Each block presented each input point described in Figure 1 to participants in a random order. A five-minute break separated the two blocks.

Data Analysis

Each experimental round contained the same game duration but employed different IAT. Thus, a different number of ships appeared in each experimental round. Therefore, it was inappropriate to compare the number of routes drawn and the total score across each experimental round as changes in IAT influenced these variables. To account for this difference, performance was measured as the percentage of the maximum possible score obtained in a game. Furthermore, engagement was calculated through two measures: human draws (HD) per ship and HD per second. When experiencing small IATs, the user may struggle to draw a route for every ship, even if the user desires to draw a manual route per ship. However, this does not mean the user is less engaged in the task than rounds where the user is physically capable of drawing a route for every ship. Therefore, it was desirable to use HD per second to measure overall engagement of a human at each IAT and ART point. However, HD per ship still proved useful for defining thresholds (i.e. we can say the human must at least engage with one in every five ships). Workload was measured using a subjective questionnaire containing three questions from NASA-TLX on a 0-20 scale. These questions were selected as previous studies found a correlation between the workload categories of temporal demand, effort, and performance with changes in IAT (Schneider et al., 2018). Workload values were standardized using min-max normalization within each participant to allow comparison across all participants. Total workload for a single Space Navigator round was calculated as the sum of the normalized workload values for each of the three workload questions.

Relationships between our independent and dependent variables were investigated using multiple linear regression. This analysis contained two steps. First, multiple regression analysis on output variables was conducted to the third order. Second, insignificant effects were removed one at a time until only significant effects remained. Regression analysis was applied for each output variable across all participants. If large participant variability caused no significance for IAT and ART across all participants, regression analysis was conducted on the mean output values for each input IAT and ART combination.

Results and Discussion

Table 1 displays correlations of IAT and ART with human-agent team performance, human engagement, and workload. Results indicated IAT strongly correlates with score ($r(8) = 0.9229$, $p = 0.0004$), engagement ($r(8) = -0.7969$, $p = 0.0642$), and workload ($r(8) = -0.9578$, $p < 0.0001$). Results also indicated that ART strongly correlates with engagement ($r(8) = 0.8481$, $p = 0.0039$). From Table 1, it becomes evident that as IAT increases, the percent of maximum possible score increases, human draws per ship increases, and workload increases. Additionally, Table 1 illustrates that as ART increases, participant engagement with the system increases. These results are consistent with data obtained in preceding research (Schneider et al., 2018).

Table 1.
Correlations between variables. Values in bold represent significant correlation at α = 0.05. Italicized data points represent significant correlation at α = 0.10.

	Avg. % Max Score	Avg. HD per Ship	Avg. HD per Sec	Avg. Std Workload
IAT (IV)	**0.9229**	*0.6385*	**-0.7969**	**-0.9578**
ART (IV)	0.0018	**0.8481**	0.3955	0.0006

Multiple linear regression analysis on the data across all participant trials indicated that there was a collective significant effect between IAT and ART on percentage of max score, F(5, 246) = 25.4565, $p < 0.0001$, $R^2 = 0.3410$. Further examination of the predictors indicated that IAT ($t = 6.14$, $p < 0.0001$, $\beta = 0.1404$), IAT to the second degree ($t = -3.42$, $p = 0.0007$, $\beta = -0.1288$), ART ($t = -3.15$, $p = 0.0018$, $\beta = -0.1263$), ART to the second degree ($t = -2.96$, $p = 0.0034$, $\beta = -0.0812$), and ART to the third degree ($t = 2.84$, $p = 0.0049$, $\beta = 0.0757$) were significant predictors in this model.

Multiple linear regression on data across all participant trials indicated there was a collective significant effect between IAT and ART on human engagement represented as human draws per ship, F(2, 249) = 16.1716, $p < 0.0001$, $R^2 = 0.1150$. Further examination of the predictors indicated that IAT ($t = 2.78$, $p = 0.0058$, $\beta = 0.0890$) and ART ($t = 4.29$, $p < 0.0001$, $\beta = 0.0746$) were significant predictors in this model.

Multiple linear regression analysis to on data across all ART and IAT combination averages indicated there was a significant effect between IAT and ART on workload, F(4, 4) = 130.1843, $p = 0.0002$, $R^2 = 0.9924$. Further examination of the predictors indicated that IAT ($t = -18.71$, $p < 0.0001$, $\beta = -0.2345$), ART ($t = 4.94$, $p = 0.0078$, $\beta = 0.0377$), ART to the second degree ($t = -3.39$, $p = 0.0275$, $\beta = -0.0265$), and the interaction of ART and IAT ($t = 3.08$, $p = 0.0370$, $\beta = 0.0535$) were significant predictors in this model

Derivation of Near-Optimal Agent Response Function

To determine the optimal ART, the regression equations derived in the previous section were applied within an optimization problem. The optimization problem was solved for the ART at each IAT value between zero and four seconds on a 0.001s interval. This optimization sought to maximize the percentage of maximum score subject to the constraints that the participant would draw at least one route for every five ships and would have a mean standardized workload between the mean, plus or minus one standard deviation of the workload from this experiment (between 0.423 and 0.561).

The optimization determined that when IAT is less than approximately 1.5s, the optimal ART is 0s. In this range, IAT is much lower than the average human response time, so the human will likely struggle to match the pace at which new tasks appear. Therefore, the human will likely require shorter ART. Once IAT is greater than 1.5s, the ART increases as IAT increases, permitting the human to take on a more involved role since they can better keep up with the slower rate at which tasks appear. As IAT approaches the average human response time, it disrupts the linear function. This permits a constant ART for IAT near the average human response time. ART then continues to increase once IAT is greater than the average human response time. Violation of the constraints specified in the function occurred at IAT greater than 3s. For this reason, ART at IAT greater than 3s was extrapolated from the function starting at IAT of 2.7s. As IAT increases from 2.7s, the human has more time to complete present tasks until the next ship arrives. Therefore, human need for agent assistance remains low at IAT levels greater than 2.7s. Optimization produces a piecewise linear function for the calculation of the optimal ART based on IAT. Equation 1 provides this piecewise linear function.

$$
\begin{aligned}
&For\ IAT < 1.485, ART = 0 \\
&For\ 1.485 \leq IAT < 2.206, ART = 3.5327 * IAT - 5.2461 \\
&For\ 2.206 \leq IAT < 2.735, ART = 2.5471 \\
&For\ IAT \geq 2.735, ART = 5.2807 * IAT - 11.8955
\end{aligned}
\tag{1}
$$

Conclusion

Results from this study indicate that IAT is strongly correlated with human-agent team performance, human engagement, and workload. Furthermore, ART is correlated with human engagement. This study produced a method for computing ART as a function of IAT. The ART function was obtained by gathering data at logical IAT and ART points and calculating which ART produced the maximum percentage of possible team score while following workload and engagement constraints. The proposed ART function will be applied in subsequent research to determine if ART calculated from IAT can effectively balance workload and engagement while maintaining equal or better performance than a constant ART agent.

Disclaimer and Acknowledgement

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Air Force, Department of Defense, nor the U.S. Government. The authors gratefully acknowledge the financial support of the Air Force Office of Scientific Research, Computational Cognition and Machine Intelligence Program.

References

Goodman, T. J., Miller, M. E., Rusnock, C. F., & Bindewald, J. M. (2017). Effects of agent timing on the human-agent team. *Cognitive Systems Research*, *46*, 40–51. https://doi.org/10.1016/j.cogsys.2017.02.007

Ishihara, S. (2012). Ishihara's design charts for colour deficiency of unlettered persons. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0,36&q=ishihara%27s+design+charts+for+colour+deficiency

Kaber, D. B., Riley, J. M., Tan, K.-W., & Endsley, M. R. (2001). On the Design of Adaptive Automation for Complex Systems. *International Journal of Cognitive Ergonomics*, *5*(1), 37–57. https://doi.org/10.1207/S15327566IJCE0501_3

Parasuraman, R. (2008). *Supporting Battle Management Command and Control: Designing Innovative Interfaces and Selecting Skilled Operators*. Fairfax, VA. Retrieved from http://www.dtic.mil/docs/citations/ADA480645

Schneider, M. F., Bragg, I. L., Henderson, J. P., & Miller, M. E. (2018). Human Engagement with Event Rate Driven Adaptation of Automated Agents. In *2018 IISE Annual Conference*. Orlando, FL.

III. Adaptive Artificial Agent Response Time Impact on Human-Agent Team Performance

Description

Chapter III includes results from a portion of a study comparing the variable agent response time teammate identified in Chapter II against a fixed agent response time teammate. The chapter highlights the importance of making agent actions predictable to their human teammates. This chapter helps to answer the third and sixth research questions specified in Chapter I.

Publication Details

The Institute of Industrial and Systems Engineering (IISE) Annual Conference and Expo accepted this article for publication. The conference will occur May 18-21, 2019 in Orlando, Florida.

Proceedings of the 2019 IISE Conference
H.E. Romeijn, A. Schaefer, R. Thomas, eds.

Adaptive Artificial Agent Response Time Impact on Human-Agent Team Performance

David J. Canzonetta and Michael E. Miller
Air Force Institute of Technology
Wright-Patterson, AFB, OH 45433, USA

Abstract

Balance of engagement and workload between members of human-agent teams has conventionally relied upon adaptive automation through dynamic function allocation. The use of adaptive agent timing may reduce mental workload present in dynamic function allocation, while still obtaining acceptable performance values. Preceding research developed an adaptive agent timing function based on the rate at which new tasks appeared to a human-agent team. Research described in these proceedings compared performance of human-agent teams using this adaptive agent response time (ART) function against performance of human-agent teams using a fixed ART function. Human-agent team performance was measured during interaction with an air-traffic-control tablet application. Results indicated that a statistical difference for human-agent team score and human engagement did not exist between the fixed and adaptive ART functions. Participants experienced higher workload with the adaptive ART function than with the fixed ART function. Feedback from participants indicated that the adaptive ART function had lower predictability than the fixed ART function. Improved predictability from an adaptive agent teammate could produce heightened results. This article explores methods for enhancing adaptive agent predictability and provides recommendations for future research.

Keywords
Human-agent teaming, agent response time, inter-arrival time, adaptive agent timing

1. Introduction
Balance of engagement and workload between members of human-agent teams has conventionally relied upon adaptive automation provided through dynamic function allocation [1]. This type of task allocation has been criticized for increasing mental workload of the operator as the human must remain cognizant of the present allocation and their duties within the present allocation [2]. The use of adaptive agent timing provides an alternate method of workload division [3], [4] and could offer a substitute for dynamic function allocation which reduces mental workload, permitting the system to balance human engagement and workload while producing acceptable human-agent team performance.

Previous research in our laboratory analyzed the effect of inter-arrival time (IAT) and agent response time (ART) on human-agent team performance, human engagement, and workload in a constant IAT environment [5]. This prior research determined that human engagement depends on ART. Furthermore, it concluded that human-agent team performance, human engagement, and workload strongly depend on IAT. This study then applied optimization techniques to calculate ART as a function of IAT for environments with variable IAT [5]. The designed ART function recommended an ART of 0 for IATs significantly shorter than the time it would take a human to respond to a new task, increasing according to the piecewise linear function in Equation (1).

$$\begin{aligned} &For\ IAT \le 1.485, ART = 0 \\ &For\ 1.485 \le IAT < 2.206, ART = 3.5327 * IAT - 5.2461 \\ &For\ 2.206 \le IAT < 2.735, ART = 2.5471 \\ &For\ IAT \ge 2.735, ART = 5.2807 * IAT - 11.8955 \end{aligned} \tag{1}$$

The study pertaining to these proceedings implemented the variable ART function along with a fixed ART function and compared the results between the two. In the current research we posit that a variable ART teammate will result in lower human workload and higher human engagement than a fixed ART teammate, while maintaining at least the same performance level.

2. Method

2.1 Participants

The experiment consisted of 16 male participants. Two participants were left-handed. The mean participant age was 27.3 and ranged from 24 to 35. All participants successfully completed the Ishihara Color Deficiency Charts prior to the experiment [6]. On average, participants spent 52.5 hours per week using a computer or similar machine.

2.2 Apparatus and Experiment

The experiment used a tablet application titled "Space Navigator." Space Navigator closely resembles commercially-available air-traffic-control games. In Space Navigator, a human and agent work together as peers to achieve the highest game score possible. The human-agent team obtains points by navigating red, blue, yellow, or green ships that spawn onto the screen to planets of their corresponding color, while obtaining randomly-appearing bonuses during their routes. Ships are removed from the screen when they arrive at their appropriate planet. Bonuses appear at a random on-screen location once every 10 seconds and remain on-screen until collected by a ship. The team loses points when two ships collide. The human can draw a path to a planet as a ship appears on the screen, but if the human does not draw a path within a specified time window, termed the agent response time (ART), the artificial agent draws a path. The artificial agent, however, only draws a straight path from the ship to its appropriate planet; it does not account for any bonuses or other ships on screen. The human can draw or redraw a route at any time. The agent cannot overwrite a human-drawn route. Participants played all games on a Microsoft Surface Pro 4 in a quiet and secluded location.

2.3 Experimental Design and Procedure

This experiment had one within-subjects, categorical independent variable: type of ART teammate, which could take values of "fixed" and "variable." If fixed, ART for the teammate remained at the average human draw time of 2.6 seconds for a round of Space Navigator. If variable, ART for the teammate was calculated as a function of current IAT using the function expressed in Equation (1).

Participants received a tutorial on the experimental environment from the research administrator, which included a demonstration that followed a narrated script. The research administrator informed participants they would work with two different teammates. However, participants received no further description of agent behavior. The participants played a single 2.5-minute practice round with each type of ART teammate. They then played an "official" fifteen minute round with each teammate. A five-minute break separated official rounds to address any participant fatigue. Half of the participants received the fixed ART teammate first, the other half received the variable ART teammate first. Participants received the same number of routing tasks across all experimental trials. Furthermore, each trial had the same duration. Participants received a workload questionnaire after each round with an agent teammate. Upon completion of the entire experiment, an additional, open-ended questionnaire asked participants whether they preferred the fixed ART teammate or the variable ART teammate.

The input IAT function to the experiment remained the same across all trials. The input IAT for all trials varied between three levels of high, medium, and low IAT. The high IAT level was defined as 3.4 seconds. IAT larger than 3.4 seconds result in situations where virtually all human operators can successfully route all ships in Space Navigator [5]. The low IAT level was defined as 1.7 seconds. IAT smaller than 1.7 seconds result in situations where virtually all human operators struggle to keep up with all ships in Space Navigator [5]. The medium IAT level was defined as 2.6 seconds, which represents the average time it takes a human operator to draw a route in Space Navigator without agent assistance [4]. Throughout the round, relaxed or rapid transitions would occur once between each IAT level at 15 and 45 seconds, respectively. Levels and transitions between levels were divided equally throughout the duration of a round.

2.4 Data Analysis
Performance was measured as the total point score obtained by the human-agent team in a single trial. Engagement was measured as the total number of human draws for a single trial. Workload was measured using NASA TLX [7] through a subjective questionnaire containing six questions on a 0-20 scale. Questions were intended to account for workload categories of mental demand, physical demand, temporal demand, performance, effort, and frustration. Additionally, the questionnaire asked participants how helpful they found their teammate on a 0-20 scale. One-factor repeated measures ANOVAs explored relationships between independent and dependent variables.

3. Results
A one-factor repeated measures ANOVA was conducted to compare the effect of teammate type on human-agent team score. There was not a significant effect of teammate type on score, $F(1, 15) = 0.036$, $p = 0.852$.

A one-factor repeated measures ANOVA was conducted to compare the effect of teammate type on human engagement, as measured by total human draws. There was not a significant effect of teammate type on human engagement, $F(1, 15) = 0.891$, $p = 0.360$.

A one-factor repeated measures ANOVA was conducted to compare the effect of teammate type on workload. There was a significant effect of teammate type on workload, $F(1, 15) = 5.156$, $p = 0.038$. Figure 1 illustrates the effect of teammate type on workload.

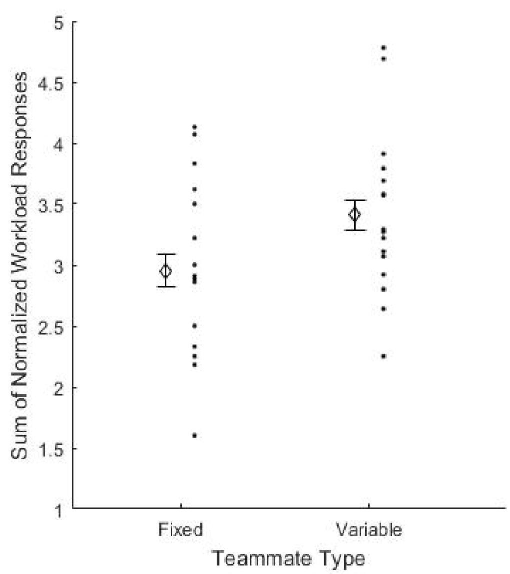

Figure 1: Graph depicting workload based on type of ART teammate. Each dot represents a workload value for a round of Space Navigator reported by a participant. Each diamond represents mean for each type of teammate. Error bars indicate standard error.

Of the 16 total participants, 13 participants indicated through open-ended feedback that they preferred the fixed ART teammate instead of the variable ART teammate. On the 0-20 scale, 7 of the 16 participants indicated they preferred the fixed ART teammate instead of the variable ART teammate. However, for 12 of the 16 participants, the difference in score on the 0-20 scale was less than four. Five of the 13 participants who preferred the fixed ART teammate actually produced a higher score with the variable ART teammate. Nine participants described the fixed teammate using a form of the words “consistent” or “predictable.”

4. Discussion

4.1 General Discussion

As hypothesized, results indicated that teams performed at least as well with the variable ART teammate as with the fixed ART teammate. However, participants reported experiencing greater workload with the variable ART teammate than with the fixed ART teammate. Furthermore, participants engaged equally with both teammates. Conceivably, the perceived unpredictability of the variable ART teammate may have contributed to higher workload reported by participants. This type of perception could also influence the willingness of participants to engage with new tasks in the experiment environment.

However, the assertion that adaptive ART can effectively balance human engagement and workload while producing acceptable human-agent team performance can still be considered. The current research tested a single function; other potential functions for adaptive ART exist which may perform better than the function applied in this research. Furthermore, the perceived enhanced predictability of the fixed ART teammate by the participants could have influenced the current results.

4.2 Participant Desire for Predictability

Participants clearly stated the reason for the fixed ART teammate preference: it appeared more "predictable" and "consistent" than a variable ART teammate. Even when producing better results with the variable ART teammate, participants generally preferred the fixed ART teammate. The desire for predictability suggests the participant inclination to trust the actions of the fixed ART teammate. This finding stands consistent with previous research that indicated predictability is a key factor in human trust of automation [8]. The increased predictability from the fixed ART teammate could also explain why participants experienced higher workload with the variable ART teammate. Absent full understanding of variable ART logic, participants likely had to dedicate cognitive resources to monitor the behavior of the variable ART teammate, thereby increasing perceived workload. Perhaps placing more emphasis on agent predictability would reduce the cognitive effort required to assess the state of the variable ART teammate and improve human trust of the variable ART teammate, thereby decreasing workload.

Methods exist that could make the variable ART teammate more predictable to the participant. For example, alteration of graphics within Space Navigator could create a more predictable variable ART teammate [9]. Currently, no visual cues exist that alert participants to the time remaining before an agent draws a route for a ship. Participants must anticipate when the variable ART teammate will draw a route. When working alongside a fixed ART teammate, participants can rely on the teammate's "rhythm." Variable ART teammate actions, however, appear more sporadic to the participants than fixed ART teammate actions. If the Space Navigator graphical interface could increase predictability of the variable ART teammate, the agent may become more desired by participants.

Additionally, a more in-depth explanation of teammate functionality to the participant during the demonstration portion of the experiment could make the variable ART teammate more predictable to the human. In this experiment, participants received information that teammates differed but had to determine this difference on their own. Future research could employ a similar experiment structure, while explicitly demonstrating to the participant that the variable ART teammate changes its response time based on IAT. This future experiment could determine if increased participant understanding of an automated teammate enhances human-agent team performance, human engagement, and workload. If a fixed and a variable ART teammate have equal predictability, potential benefits of a variable ART teammate may become apparent.

5. Conclusion

Results from this experiment did not indicate a statistical difference between human-agent team score and human engagement for a fixed and a variable Agent Response Time (ART) teammate over an entire trial. The variable ART teammate had higher perceived workload than the fixed ART teammate. Participant feedback indicated that users preferred the fixed ART teammate due to its predictability. Increased predictability from the fixed ART teammate to the participants may have influenced the results of the study. Employing methods to make the variable ART teammate as predictable as the fixed ART teammate may create a situation where the effects of adaptive agent timing can be more accurately measured.

Disclaimer and Acknowledgement

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Air Force, Department of Defense, nor the U.S. Government. The authors gratefully acknowledge the financial support of the Air Force Office of Scientific Research, Computational Cognition and Machine Intelligence Program.

References

[1] J. M. Colombi, M. E. Miller, M. Schneider, M. J. McGrogan, C. D. S. Long, and J. Plaga, "Predictive mental workload modeling for semiautonomous system design: Implications for systems of systems," *Syst. Eng.*, vol. 15, no. 4, 2012.

[2] D. B. Kaber, J. M. Riley, K.-W. Tan, and M. R. Endsley, "On the Design of Adaptive Automation for Complex Systems," *Int. J. Cogn. Ergon.*, vol. 5, no. 1, pp. 37–57, 2001.

[3] W. B. Rouse, "Human-Computer Interaction in Multitask Situations," *IEEE Trans. Syst. Man. Cybern.*, vol. 7, no. 5, pp. 384–392, May 1977.

[4] T. J. Goodman, M. E. Miller, C. F. Rusnock, and J. M. Bindewald, "Effects of agent timing on the human-agent team," *Cogn. Syst. Res.*, vol. 46, pp. 40–51, 2017.

[5] D. J. Canzonetta and M. E. Miller, "Identifying a possible function for artificial agent adaptation in variable task rate environments," in *International Symposium on Aviation Psychology*, 2019.

[6] S. Ishihara, "Ishihara's design charts for colour deficiency of unlettered persons," 2012.

[7] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Adv. Psychol.*, vol. 52, no. C, pp. 139–183, 1988.

[8] C. F. Rusnock, M. E. Miller, and J. M. Bindewald, "Observations on trust, reliance, and performance measurement in human-automation team assessment," *67th Annu. Conf. Expo Inst. Ind. Eng. 2017*, pp. 368–373, 2017.

[9] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. Van Riemsdijk, and M. Sierhuis, "Coactive Design: Designing Support for Interdependence in Joint Activity," *J. Human-Robot Interact.*, 2014.

IV. Impact of Artificial Agent Timing in Variable Inter-Arrival Time Environments

Description

Chapter IV expands upon Chapter III to include explanation of agent functionality to a group of experiment participants. Furthermore, Chapter IV analyzes human-agent team performance and human engagement at different levels of inter-arrival times. Key factors for adaptive ART function implementation are presented. Chapter IV provides insight to the third through sixth research questions listed in Chapter I.

Publication Details

The journal article contained in Chapter IV is currently pending submission to a relevant publisher. The first target is the IEEE Man Machine Systems Journal.

Impact of Artificial Agent Timing in Variable Inter-Arrival Time Environments

David J. Canzonetta, Michael E. Miller, John M. McGuirl, and Gilbert L. Peterson

***Abstract*— The current research consisted of two experiments to explore the use of adaptive agent timing as an alternative to dynamic function allocation to enhance human-agent team performance within an adaptive automation framework. The first experiment examined the effect of agent response time (ART) in a series of constant inter-arrival time (IAT) environments, each with a different IAT and ART, to understand the effect of these variables on human-agent team performance, human engagement, and workload. Results obtained from this experiment enabled the creation of an adaptive ART function based on IAT, which maximized human-agent team performance while constraining the solution to maintain desirable levels of human engagement and workload. An agent with an ART based upon the resulting adaptive ART function was compared against an agent having a fixed ART in an experiment which employed a variable IAT environment. Instructions were varied between participant groups, with one group receiving an explanation of the expected behavior of the two agents. Participant feedback conveyed a desire for improving the predictability of the agent with the adaptive ART. The results indicated that the adaptive agent timing function demonstrated the ability to reduce underload. During levels of small IAT, adaptive ART resulted in higher score and lower human engagement than fixed ART. However, participants who received explanation of agent functionality scored lower at levels of small IAT. During moderate levels of IAT, participants were more engaged with the adaptive ART teammate than the fixed ART teammate. Additionally, at the moderate level of IAT, participants who received agent instruction were more engaged with the variable ART teammate than participants who did not receive agent instruction. During levels of large IAT, adaptive ART resulted in more human engagement than fixed ART. These results help to identify key guidelines to be used in construction of adaptive ART teammates within human-agent teams.**

***Index Terms*—Human-Agent Teaming, Artificial Agent Response**

I. Introduction

Autonomous systems can enhance human ability to enrich system execution [1]. For this enrichment to occur, autonomous systems require strong human presence. Unfortunately, many of these autonomous systems limit human interaction by placing the human in a supervisory role to the agent [2], [3]. Placing the human in a supervisory role impacts the potential of the human-agent team to leverage each member's unique strengths [3]. Furthermore, though counterintuitive, humans acting in a supervisory role fall into a complacency trap where greater effort is required to maintain acceptable levels of alertness [4].

The desire to limit human interaction with artificial systems has remained the chief stance within the engineering community [5]. This viewpoint does not leverage the strengths of each team member to maximize team performance. As tasks require more induction and expertise, they require more human interaction for successful completion [6]. Conversely, as tasks become more repetitive and skill-based, they demand the precision of automation for a high rate of successful completion [6]. Tasks which vary on a skill-based to expertise scale are likely to require action from a human and an agent. This viewpoint contrasts with the viewpoint of task allocation in the traditional literature that regarded task allocation between a human and an agent as mutually exclusive [7], [8]. Based on this literature, determining an effective strategy for allocating shared human-agent tasks becomes a suitable research path.

When allocating task responsibility for shared human-agent team activities, logic dictates the avoidance of methods that inadvertently increase human workload, especially in time constrained environments which can impose levels of high human workload during periods of high task load. Many systems produce variable workload, with higher workload occurring during certain mission phases or under certain environmental conditions [9]. One approach to designing the partnership between humans and automation within this type of environment involves adaptive automation wherein the role of the automation increases to alleviate peaks in human workload [10], [11]. Traditionally, adaptive automation has employed dynamic function allocation [12]. However, this method of function allocation can increase mental workload by forcing the human to remain cognizant of their current responsibilities within present automation strategies [12]. Therefore, other task assignment methods may be worth exploring.

Previous research has suggested that the timing of agent response will likely affect the division of work between a human and an agent [3], [13]. Furthermore, recent research on the effect of agent timing in a human-agent team has demonstrated the ability of agent response timing to affect human engagement and workload within an event-driven environment [3]. Therefore, prior research suggests that one

David J. Canzonetta (e-mail: david.canzonetta.1@us.af.mil), Michael E. Miller (e-mail: Michael.Miller@afit.edu), John M. McGuirl (e-mail: jmcguirl@afit.edu), and Gilbert L. Peterson (e-mail: gpeterso@afit.edu) are with the Air Force Institute of Technology, Wright-Patterson AFB, OH 45433 USA

could institute adaptive automation through changes in agent timing as opposed to changes in function allocation [3], [14]. Changes in agent timing may shelter the user from needing to perceive and adapt their behavior to changes in the mode of automation and, therefore, might reduce the required mental workload. As suggested in the literature, changes in appropriate agent responsiveness within highly-dynamic, variable task-load environments could involve calculating agent responsiveness based on the rate at which new tasks appear [14].

Additionally, observability, predictability, and directability of agent behavior to the human can improve human-agent team performance [15], [16]. Furthermore, humans typically desire predictability of their artificial teammate actions [17]. Instruction of noticeable phenomena and directly observable graphical interface elements can increase human understanding of agent predictability [15]. However, the significance of the predictability of agent timing in a human-agent team remains unknown. Therefore, the current research will explore how the predictability of artificial agent timing affects human-agent team performance, human engagement, and workload.

The current research attempted to develop and understand the utility of an agent with variable timing. A first experiment sought to demonstrate that varying levels of arrival time between new tasks, or inter-arrival time (IAT), and agent response time (ART) across short trials could provide insight into the effects of these parameters on human-agent team performance, human engagement, and workload. These results enabled creation of a method for calculating ART as a function of IAT to generate a variable ART agent. A second experiment sought to understand the utility of the adaptive agent. It is expected that the use of the resulting variable ART teammate will improve human-agent team performance, increase human engagement, and reduce workload when compared to an agent that responds at the average human response time. Furthermore, during periods of low IAT, it is anticipated that the human-agent team will score higher and the agent will assume more of the primary task when teaming with a variable ART teammate than with a fixed ART teammate. However, during periods of high IAT, it is expected that the human-agent team will score lower, but participants will be more engaged with the variable ART teammate than with the fixed ART teammate. No significant differences in team score or participant engagement are expected when the IAT is nearly equivalent to the average human response time. In terms of predictability in agent timing, it is expected that participants who receive explanation of how the variable ART teammate operates will perform better than participants who receive no explanation of the variable ART teammate for all IAT levels, being more engaged during periods of large IAT and less engaged during periods of small IAT.

II. Experiment 1

A first experiment sought to understand the interaction of ART and IAT on human-agent team performance, human engagement, and human workload across a broad range of conditions to identify a function for calculating ART as a function of IAT.

A. Method

1) Participants

The experiment involved 14 participants (9 male and 5 female). Two participants were left-handed. The mean participant age was 25.4 and ranged from 20 to 31. One participant had previous Space Navigator experience. Participants completed the Ishihara Color Deficiency Charts prior to the experiment [18]. The results indicated that all but one participant had normal color vision. The participant with apparently anomalous color vision was allowed to complete the experiment to assess the effect of color deficiency on the play style of the game. This participant obtained the third highest recorded score, indicating their ability to successfully identify the colored items in the game. Therefore, the analysis included their data. On average, participants self-reported spending 48.7 hours per week using a computer or similar machine.

2) Apparatus and Environment

During the experiment, the participants used a tablet application titled “Space Navigator.” Space Navigator closely resembles commercially-available air-traffic-control games. Within this application, the participant and an artificial agent work together as near-peers to achieve the highest game score possible. The goal of the game is to navigate red, blue, yellow, or green ships that spawn onto the screen to planets of their corresponding color, while obtaining randomly-appearing bonuses during their routes. The human-agent team receives 100 points upon successful navigation of ships to their corresponding planet. Ships are removed from the screen when they arrive at their appropriate planet. The human-agent team loses 100 points per ship when two or more ships collide. Additionally, the human-agent team can receive 50 points for navigating ships through bonuses that appear on the screen. A bonus appears at a random on-screen location once every 10 seconds and does not disappear until collected by a ship. The human has the option of engaging in the task of drawing initial routes for the ships or acquiescing this duty to the agent. The artificial agent, however, only draws a straight path from the ship to its appropriate planet; it does not account for any bonuses or the path of any other ships on screen. Therefore, the human must perform the task of monitoring and redrawing routes to avoid collision. The human can draw or redraw a route at any time. The agent cannot overwrite any human-drawn route. Participants played all games on a Microsoft Surface Pro 4 in a quiet and secluded location.

3) Experimental Design and Procedure

The input variables to this study were inter-arrival time (IAT) and agent response time (ART). IAT is the number of seconds between the times that two subsequent ships appear. Smaller IATs indicate that more ships appear within a given time, typically leading to a higher density of ships in the game and a higher task load. ART is the time an agent waits to draw a route for a new ship. Previous research indicated that an increase in ART increased the likelihood that the human would draw an initial route for a ship and increased likelihood for a higher workload [3]. Previous research evaluated a range of IAT and ART values from 2s to 4s and 2.6s to 8.6s, respectively [14]. This research analyzed how the ratio of ART to IAT, referred to as the Adaptation Coefficient (AC), affects score, engagement, and workload [14]. However, this experiment did not include a broad enough range of conditions to shift the user's engagement with the initial route drawing task across the full range of possible conditions. Therefore, the current experimental design attempted to encapsulate a broader range of conditions.

Fig. 1 displays the IAT and ART points used in this experiment, illustrated by points with markers "x" and "o". Past studies narrowed the sampling area to boundaries and points featured in Fig. 1 [3], [14]. The dashed lines that create the top and bottom boundaries represent AC of 2.0 and 0.5, respectively. These AC were chosen because they represent locations of manageable human workload in the Space Navigator environment, as discovered in previous research [14], although human-agent team performance varied within this range. Dotted lines represent the average amount of time it takes for a human to draw a route without agent assistance, which previous research established as 2.6s [3]. Prior research defined IAT and ART boundaries shown by the dashed black lines as an area where the human continues to be human engaged while receiving adequate assistance from the adaptive automation. Previous research also demonstrated that IAT is negatively correlated with workload and ART is positively correlated with engagement [14]. Therefore, we can produce the quadrant labels in Fig. 1. When IAT is significantly less than 2.6s, the human will struggle to keep up with new tasks, thereby experiencing overload. When IAT is significantly greater than 2.6s, the human will experience a significant delay between new tasks, thereby experiencing underload. As ART decreases, the human typically draws routes slower than the agent, which could thwart the human from drawing and thereby decrease human engagement. Conversely, as ART increases, the human can draw routes faster than the agent, so one might assume that human engagement increases. However, it is also worth noting that as IAT increases, the density of ships in the environment decreases, decreasing the probability of collision. Thus, increases in IAT also reduce the urgency of route creation which might reduce human engagement.

Points marked with an "o" in Fig. 1 represent the centroid of each region within the boundaries provided by the dashed and dotted lines. Points marked with an "x" were selected to be near the boundary extremes to provide insight into human performance near these transition regions.

For each experimental session, the research administrator provided a demonstration of Space Navigator to participants from a narrated script. The participants then played three, 2.5 minute practice rounds, each with an agent teammate, to become familiar with the Space Navigator environment. Practice rounds contained slower than average IAT and ART to

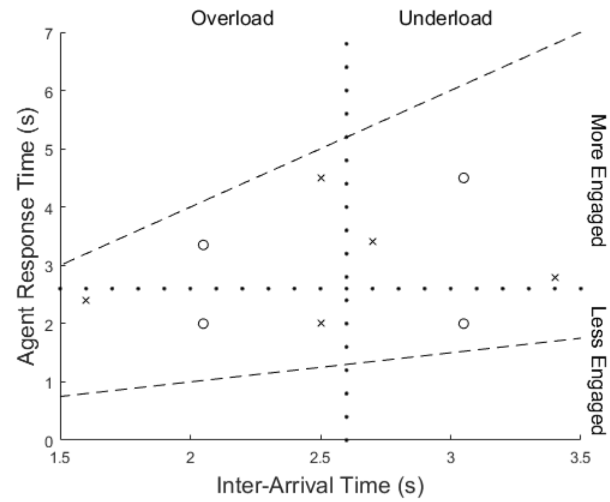

Fig. 1. Depiction of inter-arrival time (IAT) and agent response time (ART) points sampled during the current experiment (shown as x's and o's). The vertical and horizontal dotted lines indicate the average human draw time of 2.6 s. The sloped dashed lines indicate a range of values useful for human-agent teaming based on previous research. Expected generalized workload and engagement labels are associated with each quadrant in Fig 1, as defined by the dotted lines.

give participants time to understand the mechanics of the game and touchscreen response. Participants received no gameplay strategies during training.

The experimental session for each participant contained two blocks. Each block contained nine, 1.75 minute trials with a workload questionnaire administered after each trial. Game time remained constant in all trials. Each block presented each input point described in Fig. 1 in a random order. A five-minute break separated the two blocks.

4) Data Analysis

Each experimental round contained the same game duration but employed different IAT. Consequently, a different number of ships appeared in each experimental round. Therefore, it was inappropriate to compare the number of routes drawn and the total score across each experimental session as changes in IAT influenced these variables. To account for this difference, performance was measured as the percentage of the maximum possible score obtained in a game. Furthermore, engagement was calculated through two measures: total human draws (HD) per ship and total HD per second. Total HD consisted of initial draws and redraws performed by the participant. When experiencing small IATs, the user may struggle to draw a route for every ship, even if a manual route per ship is desired. However, this does not mean the user is less engaged in the task than rounds where the user is physically capable of drawing a route for every ship. Therefore, it is desirable to use HD per second to measure overall engagement of a human at each IAT and ART point. However, HD per ship is still valuable for defining thresholds (i.e. we can say the human must at least engage with one in every five ships). Workload was measured using a subjective questionnaire containing three questions from NASA-TLX on a 0-20 scale. These questions were selected as previous studies found a correlation between the workload categories of temporal demand, effort, and performance with changes in task spawn time [14] but no correlation between the remaining categories or the overall NASA TLX score. Workload values were normalized using min-max normalization within each participant to allow for comparison across all participants. Total workload for a round of Space Navigator was measured as the sum of the normalized workload values for each of the three workload questions pertaining to that round.

Data analysis included the use of simple correlation and regression analysis to understand the basic relationships between the independent variables of IAT and ART and the dependent measures discussed. Multiple linear regression explored these relationships to formulate an ART function based on IAT. This analysis contained two steps. First, multiple regression analysis on output variables was conducted to the third order. Second, insignificant effects were removed one at a time until only significant effects remained. Regression analysis was applied for each output variable across all participants. If large participant variability caused no significance for IAT and ART across all participants, regression analysis was conducted on the mean output values for each input IAT and ART combination.

B. Results

Table I displays correlations of IAT and ART with the resulting dependent variables. Results indicated IAT strongly correlates with score ($r(8) = 0.9229$, $p = 0.0004$), engagement ($r(8) = -0.7969$, $p = 0.0642$), and workload ($r(8) = -0.9578$, $p < 0.0001$). Results also indicated that ART strongly correlates with engagement ($r(8) = 0.8481$, $p = 0.0039$). Fig. 2 provides a visual representation of significant correlations noted in Table I. Each point on the individual graphs of Fig. 2 corresponds to an IAT-ART sample from Fig. 1. Error bars denote plus and minus one standard error from the mean value represented by each point.

Table I

Correlations between independent and dependent variables.

	Avg. % Max Score	Avg. HD per Ship	Avg. HD per Sec	Avg. Std Workload
IAT	**0.9229**	*0.6385*	**-0.7969**	**-0.9578**
ART	0.0018	**0.8481**	0.3955	0.0006

Values in bold represent significant correlations at α = 0.05 or less. Italicized values represent significant correlation at α = 0.10 or less.

Multiple linear regression analysis on data across all participant trials indicated there was a collective significant effect between IAT and ART on percentage of max score, $F(5, 246) = 25.4565$, $p < 0.0001$, $R^2 = 0.3410$. Further examination of the predictors indicated that IAT ($t = 6.14$, $p < 0.0001$, $\beta = 0.1404$), IAT to the second degree ($t = -3.42$, $p = 0.0007$, $\beta = -0.1288$), ART ($t = -3.15$, $p = 0.0018$, $\beta = -0.1263$), ART to the second degree ($t = -2.96$, $p = 0.0034$, $\beta = -0.0812$), and ART to the third degree ($t = 2.84$, $p = 0.0049$, $\beta = 0.0757$) were significant predictors in this model.

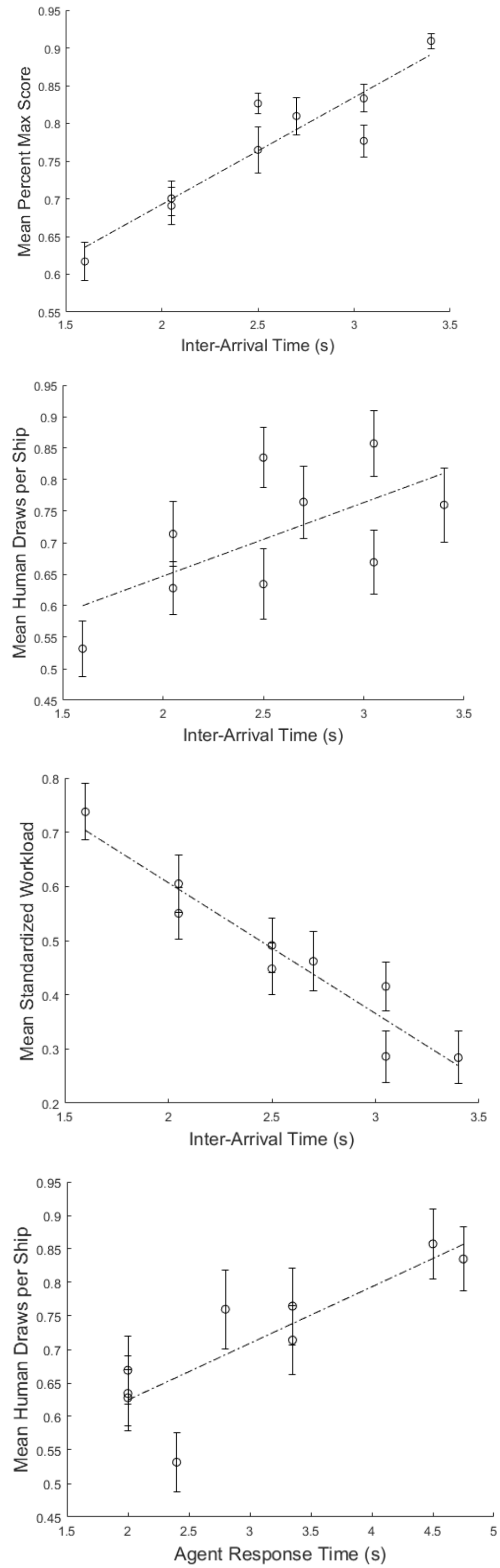

Fig. 2. From top to bottom: mean percent of maximum possible score as a function of inter-arrival time, mean human draws per ship as a function of inter-arrival time, mean workload as a function of inter-arrival time, and mean human draws per ship as a function of agent response time.

Multiple linear regression on data across all participant trials indicated there was a collective significant effect between IAT and ART on human engagement represented as human draws per ship, F(2, 249) = 16.1716, $p < 0.0001$, $R^2 = 0.1150$. Further examination of the predictors indicated that IAT ($t = 2.78$, $p = 0.0058$, $\beta = 0.0890$) and ART ($t = 4.29$, $p < 0.0001$, $\beta = 0.0746$) were significant predictors in this model.

Multiple linear regression analysis on data across all ART and IAT combination means indicated there was a significant effect between IAT and ART on workload, F(4, 4) = 130.1843, $p = 0.0002$, $R^2 = 0.9924$. Further examination of the predictors indicated that IAT ($t = -18.71$, $p < 0.0001$, $\beta = -0.2345$), ART ($t = 4.94$, $p = 0.0078$, $\beta = 0.0377$), ART to the second degree ($t = -3.39$, $p = 0.0275$, $\beta = -0.0265$), and the interaction of ART and IAT ($t = 3.08$, $p = 0.0370$, $\beta = 0.0535$) were significant predictors in this model.

III. Experiment 2

A second experiment compared human-agent team performance, human engagement, and workload between a fixed ART teammate and a variable ART teammate, as well as between participants who did or did not receive explanation on agent functionality. This experiment assumed that instructing participants on variable ART teammate functionality would permit the group receiving instruction to use the observed IAT to predict the ART, making the variable ART agent more predictable to participants receiving this instruction than to the participants who did not receive this instruction. Additionally, the experiment analyzed performance across an entire IAT function and at specific levels within the IAT function.

A. Method

1) Participants

The experiment consisted of 32 participants (29 male and 3 female). Two participants were left-handed. The mean participant age was 28.4 and ranged from 22 to 42. All participants completed the Ishihara Color Deficiency Charts prior to the experiment [18]. On average, participants self-reported spending 46.9 hours per week using a computer or similar machine.

2) Apparatus and Environment

Experiment 2 used the same apparatus as Experiment 1. Participants interacted with the Space Navigator tablet application. Experiment 2 implemented two key changes to the experiment environment. First, the environment was modified to provide a variable IAT within a single, longer trial. Second, a variable ART teammate which varied response time as a function of IAT was implemented.

Data obtained from Experiment 1 provided the formulation of the variable ART teammate implemented in Experiment 2. To determine the optimal ART, the multiple linear regression equations derived in Experiment 1 were applied within an optimization problem. The optimization problem solved for the ART at each IAT value between 0 and 4 s on a 0.001 s interval. This optimization sought to maximize the percentage of maximum score subject to the constraints that the participant would draw at least one route for every five ships and have a mean standardized workload between plus or minus one standard deviation of the mean workload identified in Experiment 1 (between 0.423 and 0.561). Fig. 3 illustrates the ART function generated from the optimization.

The linear piecewise function illustrated in Fig. 3 is shown in (1). It should be noted that the optimization failed to converge due to the constraint violations for underload when TR is greater than 3s. Therefore, it was decided to extrapolate the function found for IAT between 2.7s and 3s for all IAT greater than 3s.

The ART function in Fig. 3 demonstrates the capability of adaptive ART to influence team performance and human engagement. As IAT decreases from the average unassisted human task completion time of 2.6s, ART adjusts to boost team performance. Conversely, as IAT increases from the average unassisted human task completion time of 2.6s, ART adjusts to increase human engagement.

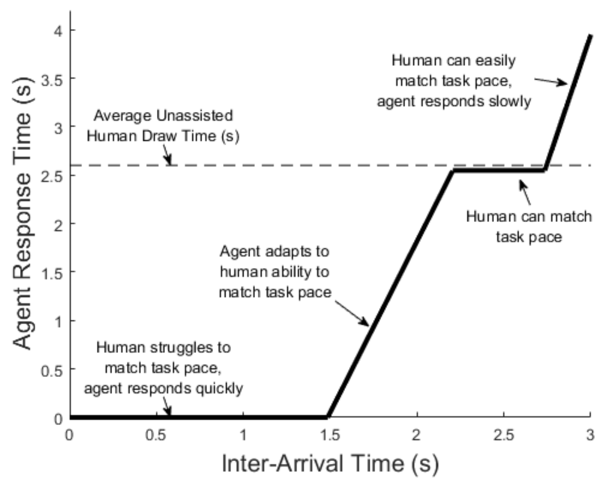

Fig. 3. Optimal ART as a function of IAT, linear regions are labeled to indicate the desired functionality of the agent within each linear region.

$$\begin{aligned}
&For\ IAT < 1.485, ART = 0 \\
&For\ 1.485 \le IAT < 2.206, ART = 3.5327 * IAT - 5.2461 \\
&For\ 2.206 \le IAT < 2.735, ART = 2.5471 \\
&For\ IAT \ge 2.735, ART = 5.2807 * IAT - 11.8955
\end{aligned} \tag{1}$$

The input IAT function to the experiment remained the same across all trials. The input IAT function for all trials varied between three levels of high, moderate, and low IAT. The high IAT level was defined as 3.4s. IATs larger than 3.4s result in situations where virtually all human operators can successfully route all ships in Space Navigator. The low IAT level was defined as 1.8s. IATs smaller than 1.8s result in situations where virtually all human operators struggle to keep up with all ships in Space Navigator. The moderate IAT level was defined as 2.6s. This is the average time it takes a human operator to draw a route in Space Navigator [3]. The input IAT function remained at each IAT level for 45s before transitioning to a different IAT level. Throughout the round, relaxed or rapid transitions would occur once between each level at 15s and 45s, respectively. Fig. 4 illustrates the resulting IAT function. Levels and transitions between levels were divided equally throughout the duration of a round.

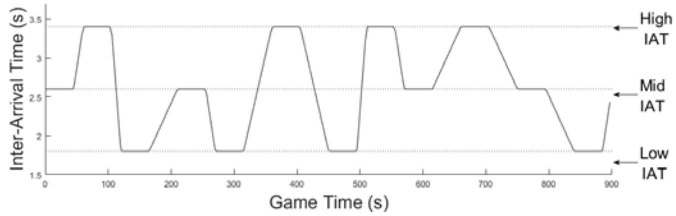

Fig. 4. Inter-arrival time (IAT) as a function of Game Time

3) Experimental Design and Procedure

The input variables to this experiment were type of ART teammate and agent explanation. Type of ART teammate is a categorical, within-subjects variable that can take values of "fixed" and "variable." If fixed, the ART for the teammate remained at the average human draw time of 2.6 seconds for a round of Space Navigator. If variable, the ART for the teammate was calculated as a function of IAT using the function shown in Fig. 3. Agent explanation is a categorical, between-subjects, variable that indicates whether the participant received instruction on how their teammate would respond prior to playing the game. A third within-subjects independent variable of IAT level was introduced when analyzing data within levels of the IAT function. IAT level is a categorical variable that represents a time within the input IAT function where IAT remains constant at 1.8, 2.6, or 3.4s.

Participants received an experiment environment tutorial from the research administrator through a demonstration that followed a narrated script. If assigned to the group receiving an explanation, participants were instructed on the functionality of their teammate. Specifically for the variable ART teammate, participants were instructed that the response time of the agent would vary as a function of IAT. This instruction permitted the participants to predict the changes in behavior of the variable ART teammate by observing the rate at which new ships were being generated in the environment. Half of the participants received this instruction.

The participants played a single 2.5 minute practice round with each of the two types of teammates. They then completed a fifteen minute experimental trial with each ART teammate. A five-minute break separated official rounds to address any participant fatigue. Half of the participants received the fixed ART teammate first, the other half received the variable ART teammate first. Workload was measured using the full NASA TLX questionnaire with a 0-20 scale. Additionally, participants were asked how helpful they found their teammate on a 0-20 scale. The participants received the questionnaire after each round with an ART teammate. Upon completion of the experiment, an open-ended questionnaire asked participants whether they preferred the fixed ART teammate or the variable ART teammate upon completion of their experiment.

4) Data Analysis

Participants received the same number of routing tasks across all experimental trials. Therefore, the output variables of team performance and human engagement did not require normalization. Team performance was measured as the total score obtained by the human-agent team in a single trial. Engagement was measured as the total number of human draws for a single trial. Workload values were normalized using min-max normalization within each participant to allow for comparison across all participants. Workload was measured as the sum of the normalized workload values for each of the six workload questions.

Performance and engagement were also compared across different IAT levels. In this type of analysis, performance was measured as a percent of maximum possible score to account for differing numbers of tasks and bonuses at the start of each IAT level. Furthermore, like analysis across different IATs in experiment 1, engagement was measured as the number of total human routes drawn per ship across the duration of an IAT level. Since workload values were obtained at the end of each trial, no workload data was available at specific IAT intervals.

Across the entire IAT function, two-factor, mixed-design ANOVAs explored relationships between independent and dependent variables. One-factor ANOVAs further investigated any interactions. When measuring performance at each IAT level, three-factor, mixed-design ANOVA explored relationships between independent and dependent variables. One-factor ANOVAs further investigated any interactions.

B. Results

First, the results across the entire IAT function for each experimental trial is presented. These results are followed by further analysis which compared the results at each IAT level.

1) Performance Across Entire IAT Function

A two-factor, mixed-design ANOVA was conducted to compare the effect of instruction and type of ART teammate on score. The ANOVA indicated that participants who received agent instruction scored lower than participants who did not receive agent instruction, $F(1, 30) = 4.416$, MSE = 26,651,406, $p = 0.044$, $\eta_p^2 = 0.128$. The ANOVA also indicated that score was not affected by the difference in fixed or variable ART teammates, $F(1, 30) = 0.768$, MSE = 2,287,656, $p = 0.388$, $\eta_p^2 = 0.025$, and the interaction between type of ART teammate and agent instruction, $F(1, 30) = 0.415$, MSE = 1,237,656, $p = 0.524$, $\eta_p^2 = 0.014$. Fig. 5 illustrates the effects of instruction and teammate type on score.

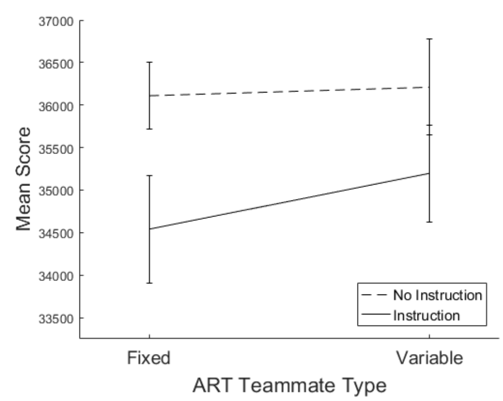

Fig. 5. Mean score for interaction of teammate type and instruction. Error bars indicate plus and minus one standard error from the mean.

A two-factor, mixed-design ANOVA was conducted to compare the effect of instruction and type of ART teammate on workload. The ANOVA indicated participants experienced greater workload with a variable ART teammate than with a fixed ART teammate, $F(1, 30) = 11.302$, MSE = 5.213, $p = 0.002$, $\eta_p^2 = 0.274$. This ANOVA also indicated that workload was not affected by agent instruction, $F(1, 30) = 0.419$, MSE = 0.362, $p = 0.523$, $\eta_p^2 = 0.014$, and interaction between type of ART teammate and agent instruction did not affect workload, $F(1, 30) = 0.426$, MSE = 0.197, $p = 0.519$, $\eta_p^2 = 0.014$. Fig. 6 illustrates the effects of instruction and teammate type on workload.

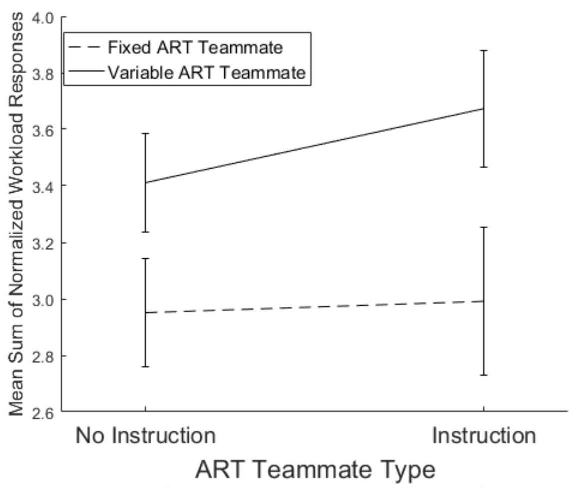

Fig. 6. Mean sum of normalized workload responses for interaction of teammate type and instruction. Error bars indicate plus and minus one standard error from the mean.

A two-factor, mixed-design ANOVA was conducted to compare the effect of instruction and type of ART teammate on engagement. The ANOVA indicated that the number of human draws was not affected by type of ART teammate ($F(1, 30) = 3.453$, MSE = 5274.4, $p = 0.073$, $\eta_p^2 = 0.103$), agent instruction ($F(1, 30) = 3.392$, MSE = 10,276.0, $p = 0.132$, $\eta_p^2 = 0.074$), or the interaction between type of ART teammate and agent instruction ($F(1, 30) = 0.831$, MSE = 1269.141, $p = 0.369$, $\eta_p^2 = 0.027$).

Of the 32 total participants, 25 participants said that they preferred the fixed ART teammate over the variable ART teammate. Thirteen of the 25 participants who preferred the fixed ART response teammate explicitly used a form of the word “predictable” and “consistent” to describe the teammate. Thirteen of the 16 participants who received no instruction and 12 of the 16 participants who received instruction indicated they preferred the fixed ART teammate.

2) Performance at Each IAT Level

A three-factor, mixed-design ANOVA was conducted to determine the effect of IAT level, instruction, and type of ART teammate on human-agent team performance. Greenhouse-Geiser correction was applied to correct the degrees of freedom for the effect of the interaction of IAT level and teammate type on engagement, which violated Mauchly’s sphericity test ($X^2(2) = 6.584$, $p = 0.037$).

The ANOVA indicated that participants who received instruction of agent functionality scored lower than participants who did not receive instruction of agent functionality, $F(1, 30) = 4.771$, MSE = 0.043, $p = 0.037$, $\eta_p^2 = 0.137$. We expected a significant interaction of IAT level and instruction, however, this interaction was not significant, $F(2, 60) = 0.711$, MSE = 0.004, $p = 0.495$, $\eta_p^2 = 0.023$. Fig. 7 illustrates mean percentage of maximum score based on the interaction of IAT level and instruction.

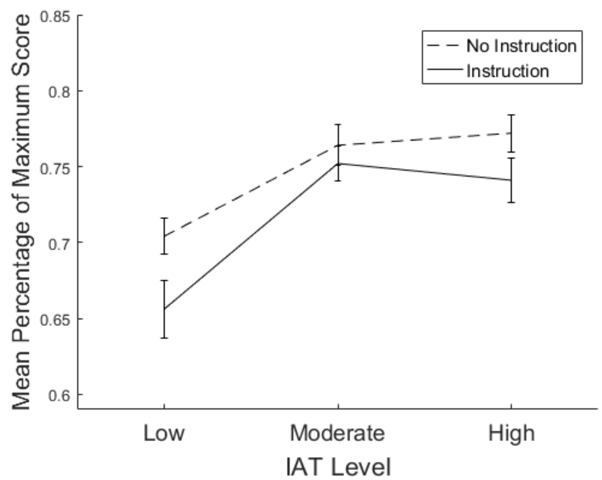

Fig. 7. Mean percentage of maximum score for interaction of instruction and IAT level. Error bars indicate plus and minus one standard error from the mean.

The ANOVA indicated that IAT level had an effect on percentage of maximum score, $F(2, 60) = 23.928$, MSE = 0.127, $p < 0.001$, $\eta_p^2 = 0.444$. Post-hoc Tukey test revealed that participants obtained a lower percentage of score at low IAT levels than at moderate or high IAT levels, $p < 0.05$. Furthermore, the ANOVA indicated that score was affected by the interaction of IAT level and teammate type, $F(1.662, 49.871) = 13.600$, MSE = 0.057, $p < 0.001$, $\eta_p^2 = 0.312$. To further investigate this interaction, a one-factor, repeated measures ANOVA was conducted to compare the effect of teammate type on score for each IAT level. For moderate IAT levels, the ANOVA indicated there was not a significant effect of teammate type on score, $F(1, 31) = 0.518$, MSE = 0.002, $p = 0.477$, $\eta_p^2 = 0.016$. For high IAT levels, the ANOVA indicated that participants scored higher with a fixed ART teammate than with a variable ART teammate, $F(1, 31) = 4.150$, MSE = 0.024, $p = 0.050$, $\eta_p^2 = 0.118$. For low IAT levels, the ANOVA indicated that participants scored higher with a variable ART teammate than with a fixed ART teammate, $F(1, 31) = 18.151$, MSE = 0.070, $p < 0.001$, $\eta_p^2 = 0.369$. Mean percentage of maximum score based on IAT level and instruction is illustrated in Fig. 8.

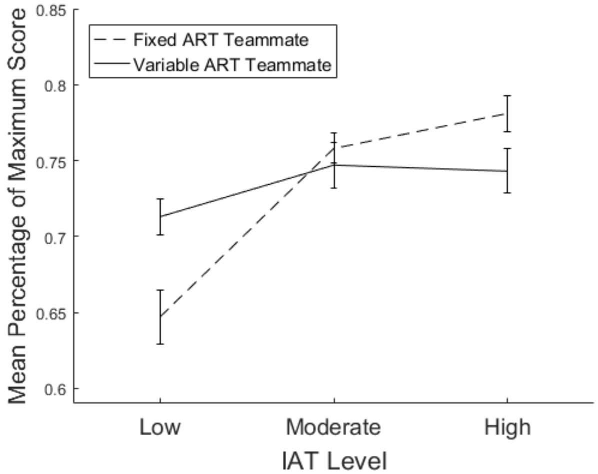

Fig. 8. Mean percentage of maximum score for interaction of teammate type and IAT level. Error bars indicate plus and minus one standard error from the mean.

Additionally, the ANOVA indicated that score was not affected by teammate type, $F(1, 30) = 0.227$, MSE = 0.001, $p = 0.637$, $\eta_p^2 = 0.008$, interaction of teammate type and instruction, $F(1, 30) = 0.057$, MSE = 0.000, $p = 0.814$, $\eta_p^2 = 0.002$, interaction of IAT level and instruction, $F(2, 60) = 0.711$, MSE = 0.004, $p = 0.495$, $\eta_p^2 = 0.023$, and interaction of teammate type, IAT level, and instruction, $F(2, 60) = 2.528$, MSE = 0.009, $p = 0.088$, $\eta_p^2 = 0.078$.

3) Engagement at Each IAT Level

A three-factor, mixed-design ANOVA was conducted to determine the effect of IAT level, instruction, and type of ART teammate on human engagement. Greenhouse-Geiser correction was applied to correct the degrees of freedom for engagement based on IAT level, which violated Mauchly's sphericity test ($X^2(2) = 9.506$, $p = 0.009$).

The ANOVA indicated that participants were more engaged with a variable ART teammate than with a fixed ART teammate, $F(1, 30) = 7.639$, MSE = 0.333, $p = 0.010$, $\eta_p^2 = 0.203$. It also indicated that IAT level had an effect on human engagement, $F(1.563, 46.894) = 132.761$, MSE = 2.113, $p < 0.001$. A post-hoc Tukey test revealed that participants drew the least number of routes per ship at the low IAT level and the most number of routes per ship at the high IAT level, $p < 0.05$.

Furthermore, the ANOVA indicated that engagement was affected by the interaction of teammate type and IAT level, $F(2, 60) = 62.040$, MSE = 0.665, $p < 0.001$, $\eta_p^2 = 0.674$. To further investigate this interaction, a one-factor, repeated measures ANOVA was conducted to compare the effect of teammate type on engagement for each IAT level. The one-factor ANOVA indicated that that participants were more engaged with the variable ART teammate than with the fixed ART teammate at moderate IAT levels, $F(1, 31) = 4.561$, MSE = 0.077, $p = 0.041$, $\eta_p^2 = 0.128$, and high IAT levels, $F(1, 31) = 35.680$, MSE = 1.382, $p < 0.001$, $\eta_p^2 = 0.535$. However, the one-factor ANOVA also indicated that participants were less engaged with the variable ART teammate than with the fixed ART teammate at low IAT levels, $F(1, 31) = 19.734$, MSE = 0.205, $p < 0.001$, $\eta_p^2 = 0.389$. Fig/ 9 illustrates mean engagement based on IAT level and teammate type.

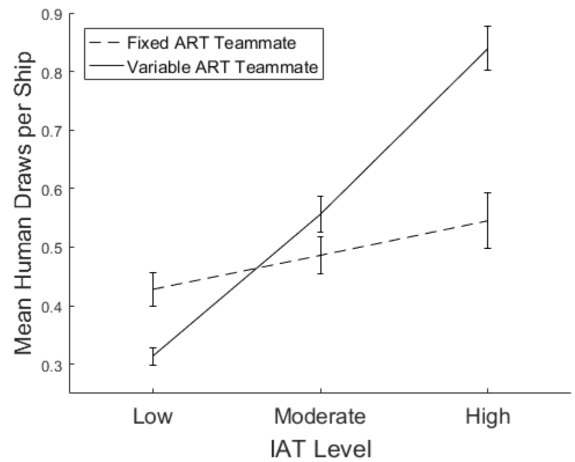

Fig. 9. Mean human draws per ship for interaction of teammate type and IAT level. Error bars indicate plus and minus one standard error from the mean.

Additionally, the ANOVA indicated that engagement was not affected by instruction, $F(1, 30) = 2.562$, MSE = 0.267, $p = 0.120$, $\eta_p^2 = 0.079$, the interaction of teammate type and instruction, $F(1, 30) = 1.382$, MSE = 0.060, $p = 0.249$, $\eta_p^2 = 0.044$, the interaction of IAT level and instruction, $F(2, 60) = 0.933$, MSE = 0.012, $p = 0.399$, $\eta_p^2 = 0.030$, and the interaction of teammate type, IAT level, and instruction, $F(2, 60) = 1.461$, MSE = 0.016, $p = 0.240$, $\eta_p^2 = 0.046$.

IV. Discussion

In a broader sense, IAT is a mechanism to control total task load. To some degree, ART represents the amount of task load assumed by the agent. Generally, as the task load increases, the agent should become more responsive and as the task load decreases, the agent should become less responsive. This enables the agent to assume more or less of the task load per unit time.

Varying levels of task load and agent responsiveness across short trials provided insight into their effect on human-agent team performance, human engagement, and workload. Results from Experiment 1 demonstrate that task load strongly correlates with score, human engagement, and workload. Furthermore, agent responsiveness strongly correlates with human engagement. Table I and Fig. 2 reveal that as IAT creates a task load decrease, team percent of maximum possible score increases, human engagement increases, and human workload decreases. Additionally, as the agent becomes less responsive, participant engagement with the system increases. Agent responsiveness did not appear to influence score or workload. Prior research supports these results [14].

Data obtained in Experiment 1 contributed to the development of an ART function calculated based on current IAT. The data created an optimization problem that maximized participant score while maintaining adequate engagement (i.e. human engaging with at least 20% of ships) and workload (i.e.

within one standard deviation of standardized mean workload across participants). This optimization was performed for IAT between 0s and 4s, resulting in a piecewise linear function for ART and providing an objective function for the variable ART teammate in Experiment 2. The resulting agent responded immediately for IAT less than 1.5s, then increased linearly to an ART of 2.5s at an IAT of 2.2s. Between IAT of 2.2s and 2.7s, the human can likely match the arrival of new ships because the ART remains fixed at 2.5s, which nearly equals the average human reaction time. This ART permits the human more flexibility to monitor and redraw routes to avoid impending collisions. For larger IATs, the ART increases at a rate of 5.3s per second of IAT, theoretically leaving the human to draw most of the routes. Note that the optimization is constrained by the engagement or workload limits for IATs greater than 3s, implying that the design is sacrificing peak performance in order to keep the human engaged in the task.

Contrary to the hypothesis, no significant difference existed between score and type of agent teammate across the entire IAT function. Perhaps the fixed ART teammate, which consistently assumed the same amount of task load, made it easier for the participants to work alongside than originally anticipated. However, during periods of high task load, participants scored significantly higher with the variable ART teammate than with the fixed ART teammate. Furthermore, and as hypothesized, type of agent teammate did not affect score during periods of moderate task load. This result likely exists because participants experience equal assumption of task load by the agent for both types of teammates at moderate task load levels. As hypothesized, teams scored higher with the fixed ART teammate than with the variable ART teammate during low task load levels. During periods of low task load, participants put forth effort to determine the precise moment at which their teammate would assume the task. This effort may have existed even if participants understood teammate functionality. Conceivably, the effort required to predict when the teammate would draw a route contributed to the variation in score between the fixed and variable ART teammate for low task load levels.

Also differing from the hypothesis, humans engaged equally with the fixed and variable ART teammates across the entire IAT function. This result suggests the influence of the variable ART teammate at differing levels of task load. For example, as hypothesized, at high task load levels, participants felt more engaged with the fixed ART teammate than with the variable ART teammate because the timing adjustment of the variable ART teammate enabled it to respond to tasks faster than the participants could generally draw an initial route. Conversely, and as posited, the participants experienced more engagement with the variable ART teammate than the fixed ART teammate at low task load levels because the adjusted delayed response of the variable ART teammate encouraged participants to draw routes manually. Since the variable ART teammate contributes to greater human engagement than the fixed ART teammate at low task load levels, but less at high task load levels, the total engagement effectively evens out, which helps to explain why no significant difference exists in engagement with both types of teammates across the entire IAT function.

Human engagement at moderate task load levels, however, indicated that participants drew significantly more routes with the variable ART teammate than with the fixed ART teammate. This finding is particularly interesting because ART was essentially equivalent at this level for both types of teammates. This demonstrates that something about the variable ART teammate influenced the participants to draw more routes at this level. Perhaps the slower response of the variable ART teammate at low task load levels subconsciously made participants more involved during periods of moderate task load. Previous research enables the assertion that humans abdicate responsibility to an agent when the agent rapidly presents its decision [19].

Contrary to the hypothesis, participants experienced higher workload with the variable ART teammate than with the fixed ART teammate across the entire IAT function. The perceived unpredictability of the variable ART teammate could contribute to a workload increase. Furthermore, the fixed ART teammate could have assumed more tasks during periods of human underload, therefore resulting in lower perceived workload.

In terms of human-agent team performance from the scope of the entire IAT function, results indicated that participants who received instruction on teammate functionality scored significantly lower than participants who did not receive instruction. This represents a potentially counterintuitive result, as one would likely predict that instruction of agent functionality would increase score. Possibly, this counterintuitive result arises from the difference between the skill levels of participants. For example, the three highest scores came from individuals receiving no instruction on agent functionality. Perhaps those individuals would have performed better with deeper understanding of the workings of their teammate. Additionally, the introduction of agent instruction could create an additional stressor for participants. The high task load level, where a new task appeared every 1.8 seconds, also demonstrated this effect. At this task load level, participants who received agent instruction scored significantly lower than participants who did not receive agent instruction. However, while virtually no difference existed between the mean score of the fixed and the variable agent for the participants who did not receive instruction, participants who received agent instruction scored significantly higher with the variable ART teammate than with fixed ART teammate.

No significant difference existed between teams that did and did not receive instruction on agent functionality at low and moderate task load levels. This makes sense at low task load levels because the slow rate of new tasks enables an acceptable score regardless of any kind of instruction. The reason for the significant difference at moderate task load levels remains unclear. Perhaps participants clearly understood the function of the agent at this task load level without any kind of instruction.

Contrary to the hypothesis, instruction of agent functionality did not affect engagement across the entire IAT function or at low or high task load levels. Instruction could have affected the perception of the variable ART teammate to participants, while not affecting participant behavior at these levels. However, when tasks appeared at a rate of 2.6 seconds, which represents the average human time to respond to a task, participants who received agent instruction experienced more engagement with the variable ART teammate than participants who did not receive agent instruction. This indicates instructing participants on the nuances of their teammate actions influenced their engagement within the game.

Workload was not affected by agent instruction, disagreeing with the hypothesis. Potentially, the explanation of teammate functionality did inadequately addressed predictability issues with the variable ART teammate. Even with the explanation of agent functionality, participants still did not know the exact time when the variable ART teammate would draw a route, they only had an indirect indication of this time (i.e., perceived IAT).

Additionally, participants clearly stated reason for the fixed ART teammate preference: it was more "predictable" and "consistent" to the participants than a variable ART teammate. Participants generally preferred the fixed ART teammate even when producing better results with the variable ART teammate. The desire for predictability indicated the participant inclination to trust the actions of the artificial teammate. Previous research supports this finding that indicated trust of automation depends on predictability [20]. The clear desire for agent predictability over an IAT function with equal amounts of relaxed and rapid transitions between IAT levels suggests the presence of a predictability need regardless of changes in task load. However, the possibility also exists that in more natural conditions, where the change in task load is likely more gradual than in the current experiment, there will be less of a need to predict the agent's response time than in the current experimental arrangement.

V. Conclusion

This research identified key guidelines to be used in development of adaptive agent response time (ART) teammates within human-agent teams. As exploration of effective adaptive ART continues, these takeaways merit consideration. First, as task load increases, an agent that responds quickly to assume more task load can boost human-agent team performance. Second, as task load decreases, a delayed agent response can maintain acceptable human engagement levels. Third, predictability of an agent is a trait of automation sought by humans. While previous research determined humans desire predictability from artificial teammate actions [17], the current research further suggests that participants are specifically seeking predictability of the agent's timing in environments where the agent's timing is likely to vary. As a result, it is suggested the design of human-agent teams must consider agent timing and agent function. Further, this research developed a method to create a variable agent timing function that results in desired general behavior. Optimization techniques to maximize human-agent team performance enable flexible constraints that can be adjusted to manipulate human behavior.

Disclaimer and Acknowledgment

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Air Force, Department of Defense, nor the U.S. Government. The authors gratefully acknowledge the financial support of the Air Force Office of Scientific Research, Computational Cognition and Machine Intelligence Program.

References

[1] Department of Defense: Defense Science Board, "Task Force Report: The role of autonomy in DoD Systems," 2012.

[2] D. J. Bruemmer, J. L. Marble, and D. D. Dudenhoeffer, "Mutual initiative in human-machine teams," in *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, 2002, pp. 22–30.

[3] T. J. Goodman, M. E. Miller, C. F. Rusnock, and J. M. Bindewald, "Effects of agent timing on the human-agent team," *Cogn. Syst. Res.*, vol. 46, pp. 40–51, 2017.

[4] R. Parasuraman, "Supporting Battle Management Command and Control: Designing Innovative Interfaces and Selecting Skilled Operators," Fairfax, VA, 2008.

[5] L. J. Bannon, "From Human Factors to Human Actors: The Role of Psychology and Human-Computer Interaction Studies in System Design," *Readings Human–Computer Interact.*, pp. 205–214, Jan. 1995.

[6] M. M. Cummings, "Man versus machine or man + machine?," *IEEE Intell. Syst.*, vol. 29, no. 5, pp. 62–69, 2014.

[7] P. Fitts, "Human engineering for an effective air navigation and traffic control system," *Ohio State Univ.*, 1951.

[8] H. E. Price, "The Allocation of Functions in Systems," *Hum. Factors*, vol. 27, no. 1, pp. 33–45, 1985.

[9] K. M. Feigh, M. C. Dorneich, and C. C. Hayes, "Toward a Characterization of Adaptive Systems: A Framework for Researchers and System Designers," *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 54, no. 6, pp. 1008–1024, 2012.

[10] M. W. Scerbo, "Theoretical Perspectives on Adaptive Automation," in *Automation and Human Performance*, Routledge, 2018, pp. 57–84.

[11] C. E. Billings, *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers, 1997.

[12] D. B. Kaber, J. M. Riley, K.-W. Tan, and M. R. Endsley, "On the Design of Adaptive Automation for Complex Systems," *Int. J. Cogn. Ergon.*, vol. 5, no. 1, pp. 37–57, 2001.

[13] W. B. Rouse, "Human-Computer Interaction in Multitask Situations," *IEEE Trans. Syst. Man. Cybern.*, vol. 7, no. 5, pp. 384–392, May 1977.

[14] M. F. Schneider, I. L. Bragg, J. P. Henderson, and M. E. Miller, "Human Engagement with Event Rate Driven Adaptation of Automated Agents," in *2018 IISE Annual Conference*, 2018.

[15] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. Van Riemsdijk, and M. Sierhuis, "Coactive Design: Designing Support for Interdependence in Joint Activity," *J. Human-Robot Interact.*, 2014.

[16] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, "Ten challenges for making automation a 'team player' in joint human-agent activity," *IEEE Intell. Syst.*, vol. 19, no. 6, pp. 91–95, 2004.

[17] J. M. Bindewald, M. E. Miller, and G. L. Peterson, "Creating Effective Automation to Maintain Explicit User Engagement," *Int. J. Hum. Comput. Stud.*, 2019.

[18] S. Ishihara, "Ishihara's design charts for colour deficiency of unlettered persons," 2012.

[19] C. Layton, P. J. Smith, and E. McCoy, "Design of a Cooperative Problem-Solving System for En-Route Flight Planning: An Empirical Evaluation," *Hum. Factors*, vol. 36, no. 1, pp. 94–119, 1994.

[20] C. F. Rusnock, M. E. Miller, and J. M. Bindewald, "Observations on Trust , Reliance , and Performance Measurement in Human-Automation Team Assessment," in *Proceedings of the 2017 Industrial and Systems Engineering Conference*, 2017.

V. Conclusions and Recommendations

Chapter Overview

This chapter answers the initial research questions. It also provides recommendations for future research in human-agent teaming within the Space Navigator testing environment. Lastly, it states the significance of research presented in this thesis.

Evaluation of Research Questions

This section re-states initial research questions. A discussion expands upon each research question.

1. *How do we determine a method for the effective timing of an artificial agent within a variable IAT environment?*

Data across different IAT and ART points produced regression functions for each response variable. Optimization techniques applied to the regression functions shaped an ART function that maximized the team performance regression function while adhering to constraints defined by the engagement and workload regression functions. Conceivably, other techniques for determining the most effective ART exist, but optimization sufficed for available data.

2. *How do IAT and ART relate to human-agent team performance, human engagement, and workload?*

Research indicated that IAT is strongly correlated with human-agent team performance, human engagement, and workload. It also indicated that ART is strongly correlated with human engagement. The data obtained in this research remained consistent with data obtained in previous research (Schneider et al., 2018).

3. *How does a variable ART teammate, as compared to a fixed ART teammate, affect human-agent team performance, human engagement, and human workload?*

Across an entire IAT function, differences between the variable and fixed ART teammates used in this research did not affect human-agent team performance or human engagement. However, the variable ART teammate used in this research created more workload for the human than the fixed ART teammate. The perceived unpredictability of the variable ART teammate may have contributed to the workload increase. Furthermore, the variable ART teammate could have assumed fewer tasks during periods of human underload, and therefore affected the perception of the teammate by the human in a manner that increased workload during low workload conditions, resulting in potentially desirable changes in workload.

Across IAT levels, teammate type did not affect score. However, participants were more engaged with a variable ART teammate than with a fixed ART teammate across the constant IAT levels, which contrasts with type of teammate effect on engagement obtained across an entire IAT function. The omission of IAT transition periods within the IAT function while conducting analysis at IAT levels is the likely reason for this difference. Data at each IAT level demonstrates that well-defined differences in IAT contribute to a situation where teammate type can influence human engagement.

4. *How does IAT level affect human-agent team performance and human engagement?*

This research determined that participants obtained the highest percentage of maximum possible score at a high IAT level and the lowest percentage of maximum possible score at the low IAT level. Results at each IAT level in Experiment 2 reinforced the relationship established in Experiment 1 that IAT is strongly correlated with human-agent team performance. Percentage of maximum possible score is expected to increase as IAT increases.

Furthermore, this research determined that participants drew the least number of routes per ships at the low IAT level and the greatest number of routes per ship at the high IAT level. Again, results from Experiment 2 fit the relationship established in Experiment 1 that IAT is strongly correlated with human engagement. Human engagement is expected to increase as IAT increases.

5. *How does explanation of agent functionality to a human affect human-agent team performance, human engagement, and human workload?*

Across an entire input IAT function, participants who received agent instruction scored lower than participants who did not receive agent instruction. This result was consistent for both the fixed and variable agent conditions, indicating that this difference was likely due to a difference in skill level between the participant groups rather than the effect of instruction. For example, the three highest scores came from individuals receiving no instruction on agent functionality. Perhaps those individuals would have performed better with deeper understanding of their teammate functionality. Additionally, the introduction of agent instruction could create an additional stressor for participants that they must remain aware of throughout the game. At the low IAT level, like the results from the entire input IAT function, participants who received instruction on agent functionality scored significantly lower than participants who did not receive agent instruction.

Instruction on agent functionality did not affect score at moderate and high IAT levels. This makes sense at high IAT levels because the slow rate of new tasks enables an acceptable score regardless of any kind of instruction. It is unclear why there was no significant difference at moderate IAT levels. Perhaps the function of the agent is clear enough at this level without instruction.

Instruction of agent functionality did not affect human engagement across the entire IAT function or at low or high IAT levels. Instruction could have affected the perception that individuals had of the variable ART teammate, while not affecting their behavior at these levels. However, when tasks appeared at a rate of 2.6 seconds, which represents the average human time to respond to a task, participants who received agent instruction were more engaged with the variable ART teammate than participants who did not receive instruction. This indicates instructing participants on nuances of their teammate influenced their actions within the game.

Workload was not affected by agent instruction. Conceivably, explanation of teammate functionality did not adequately address predictability issues with the variable ART teammate. Even with the explanation of agent functionality, participants still did not know the exact time at which the variable ART teammate would draw a route, they only had a better idea of the concept.

6. *How do humans view a variable ART teammate compared to a fixed ART teammate in terms of predictability?*

Seventy-eight percent of participants indicated they preferred the fixed ART teammate instead of the variable ART teammate. Fifty-two percent of those participants explicitly used a form of the words “consistent” or predictable” to describe the fixed ART teammate. Perhaps even more surprising, ten participants who scored higher with the variable ART teammate stated they preferred the fixed ART teammate for predictability advantages. Clearly, participants in this study valued predictability. This suggests that creation of automated teammates in future research should employ methods that increase predictability of agent actions.

Recommendations for Future Research

This research has uncovered one potential function for a variable ART teammate. Conceivably, other functions exist that might produce better human-agent team performance while improving engagement and workload.

Furthermore, the variable ART function identified in the first experiment is generalized across all potential participants as a “one size fits all” solution. However, feedback from participants suggested that they employed different strategies when playing Space Navigator. Therefore, it may prove worthwhile to create a customized variable ART teammate for a single user that leverages specific strengths and downplays any weaknesses of the user. If designing for individual participants proves too complicated, it may be easier to develop variable ART teammates around common play styles and fit new participants to a teammate based on their play style. Possibly, some type of individualization of the variable ART teammate will result in better human-agent team performance than a generalized variable ART teammate.

It could also prove interesting to quantify the impact of predictability on human-agent team performance. Improved graphical methods within the Space Navigator environment have the potential to increase variable ART teammate predictability to the human. For example, a ship in Space Navigator could visually cue the human to the precise moment their teammate will draw a route. Comprehension of agent predictability impact on human-agent teams could influence the creation of a variable ART teammate that maximizes human-agent team performance.

Possibly, dynamic variables within the scope of a game could influence optimal ART. For example, the number of open tasks (i.e. number of ships on screen) could be a better indicator of task load than IAT. Therefore, future research might explore alternative metrics to trigger changes in ART. The Space Navigator tool does not currently possess the capability to dynamically adjust its calculation of ART. The addition of this capability would allow for an enhanced study of the impact of changing variables on optimal ART.

Additionally, this thesis postulated that adaptive agent timing could serve as a substitute for dynamic function allocation by reducing the demonstrated workload penalty incurred with dynamic function allocation. However, an adaptive ART teammate created more workload for participants than a fixed ART teammate. Since the implementation of dynamic function allocation was not an input condition to this thesis, it remains unknown how the increase in workload from ART compares to the increase in workload created by dynamic function allocation. Future research could analyze workload effects created by dynamic function allocation against workload effects created by adaptive agent timing.

Significance of Research

In a broader sense, inter-arrival time (IAT) is a mechanism to control total task load. To some degree, agent response time (ART) represents the amount of task load assumed by the agent. Generally, as task load increases, the agent should become more responsive and as task load decreases, the agent should become less responsive. This enables the agent to assume more or less of the task load per unit time.

This research identified key guidelines to be used in the development of adaptive ART teammates within human-agent teams. As exploration of effective adaptive ART continues, these takeaways merit consideration. Three key ideas were established:

1. As task load increases, an agent that is quick to assume more task load can boost human-agent team performance.
2. As task load decreases, an agent that is slow to assume more task load can maintain acceptable human engagement levels.
3. Predictability of agent action is a trait of automation sought by humans. While previous research determined humans desire predictability from automated teammate actions, this thesis further suggests that participants are specifically seeking predictability from their adaptive ART teammates in real-time environments where task load is continuously changing.

Furthermore, this research developed a method to create a variable agent timing function that results in desired general behavior. Optimization techniques to maximize human-agent team performance enable flexible constraints that can be adjusted in future research to manipulate human behavior within human-agent teams.

This research fits into the larger scope of identifying effective strategies for teaming a human with an artificial agent. As autonomous systems become more prevalent in USAF operations, researchers must answer investigative questions such as the ones presented in this thesis to effectively pair humans and automated agents. The takeaways listed above enable further research of the effectiveness of using adaptive agent timing as opposed to dynamic function allocation. Ultimately, this research avenue has potential to effect teaming of humans with automated agents in the USAF.

Appendix A: Pre-Experiment Questionnaire

Pre-Test Questionnaire

Participant ID: _______

Age: ______

Gender (Circle): Male Female

Highest Degree Obtained: GED/High School Diploma Bachelor's Master's PhD

Handedness: Right Left Ambidextrous

Are you currently a member of the U.S Armed Forces? Yes No

How many hours per week do you spend on a computer-based machine (PC, tablet, phone, video game console, etc.)? ______________

Approximately how many hours of sleep did you receive last night? __________

Do you have any previous experience with the Space Navigator application? Yes No

If yes, approximately how many hours have you spent interacting with the application? _______

Appendix B: Experiment 1 Post-Round Questionnaire

Round ID_____

How hurried or rushed was the pace of the task?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How hard did you have to work to accomplish your level of performance?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How successful were you in accomplishing what you were asked to do?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How helpful did you find your automated teammate?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Not Helpful Very Helpful

What about your automated teammate did you find helpful/not helpful?

Appendix C: Experiment 2 Post-Round Questionnaire

Round ID_____

How mentally demanding was the task?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How physically demanding was the task?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How hurried or rushed was the pace of the task?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How successful were you in accomplishing what you were asked to do?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Success Failure

How hard did you have to work to accomplish your level of performance?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How insecure, discouraged, irritated, stressed, and annoyed were you?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Very Low Very High

How helpful did you find your automated teammate?

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Not Helpful Very Helpful

What about your automated teammate did you find helpful/not helpful?

What strategy did you employ to obtain a high score?

Appendix D: Experiment 2 Post-Experiment Questionnaire

Post Experiment Questionnaire

Did you find your first or second teammate to be more helpful? Why?

____

____

____

Did you find your first or second teammate more effective in maintaining a constant level of effort? Why?

____

____

Appendix E: Proctor Script

<When it is time to begin the experiment, Proctor ensures tutorial game 1 is pre-loaded into Space Navigator>

<*Proctor places tablet in front of participant with Main Menu Screen open*>

This experiment utilizes a tablet based application game called “Space Navigator.” The goal of Space Navigator is to achieve the highest game score possible. I will now demonstrate to you how to play this game.

<*Proctor presses play*>

These moving objects are ships. <*Proctor points to ships*>. These large stationary objects that are red, blue, green, or yellow are planets <*Proctor points to planets*>. Smaller stationary brown objects with white orbs are bonuses. <*Proctor points to bonuses*>. The object of the game is to obtain the most points possible. Points are obtained by getting ships to intersect with planets of their corresponding color or ships of any color to intersect with bonuses. Ships are moved by drawing a path with your finger. <*Proctor demonstrates how to draw with finger or stylus*>. Paths can then be drawn to the corresponding planet. *<Proctor draws path from ship to planet, waits until intersection>*. Routes can be redrawn for a ship as many times as you desire. Notice how the intersection of a planet and ship of the same color results in 100 points. This same process would work for bonuses and result in 50 points <*Proctor draws path from ship to bonus*>. If ships collide however, the result is a loss of 200 points, 100 points for each ship. <*Proctor demonstrate two ships colliding*>. Furthermore, notice how ships will not explode when intersecting planets of opposite colors. <*Proctor demonstrates ship and planet of different colors colliding.*> Moreover, when a ship flies off screen without a route, no points are lost, but that specific ship is no longer a part of the game. <*Proctor demonstrates ship flying off screen.*> Notice that your current score is visible in the top left corner of the screen. Time remaining is in the top right corner of the screen.

<Proctor loads tutorial game 2 containing a straight-line agent into Space Navigator and presses play, then waits for an agent to draw a path>

Now, you aren’t playing this game alone! You will have an automated teammate that is designed to assist in your performance throughout the game. Notice how if you do not input a path, your teammate draws a straight-line to a planet for you. Notice that this line does not navigate towards any bonuses or around any ships. Furthermore, your teammate only draws a line once per ship. Once you draw or redraw a path, your teammate will no longer provide assistance for that ship. This concludes the demonstration of Space Navigator, do you currently have any questions?

For this experiment, you will have two types of teammates assisting you.** You will have the opportunity to play a practice round with each teammate. Upon completion of the practice rounds, you will play an official round with each teammate. The order in which you play with each teammate is random. Upon completion of each round, please hand the tablet back to me. Upon completion of the practice games, we will play two official rounds. To reiterate, each round will utilize a teammate that implements a different strategy to assist you. I will let you know when the official rounds are about to begin. Before we go any further, are there any questions that you have?

<Proctor answers any questions and shows NASA-TLX to participant>

This questionnaire will be given upon completion of each round of Space Navigator. Please follow the directions for each question and circle the number you feel is appropriate. *<Proctor notates proper way to mark NASA-TLX>*. We are now going to begin the official portion of the experiment. Please hand the tablet back to me when the round ends.

<Proctor conducts experiments>

<Experiment concludes>

This experiment has concluded. Thank you for your participation! Results of the study will be published upon research completion. If you are interested in finding out the results, I can notify you upon research completion. Furthermore, you can stop by this lab to check high scores as often as you would like. For reference, your ID is _______.

*** <For Experiment 2, refer to Appendix F and continue there before returning to this script.>*

Appendix F: Proctor Explanation of Automated Teammate Functionality

<u>Information provided to participant before introducing teammate in the Space Nav tutorial:</u>

You will be working with two teammates. One teammate will respond based on the rate at which ships appear. In other words, it will respond more rapidly as ships appear at a faster rate and less rapidly as ships appear at a slower rate. The other teammate will respond at the same rate throughout the entire game. In other words, it will maintain a constant response time as ships appear both more and less rapidly on-screen.

<u>Information provided to participant before fixed ART teammate practice and official rounds:</u>

You will now be working with your constant response teammate. It will respond at the same rate throughout the entire round. In other words, it will maintain a constant response time as ships appear both more and less rapidly on-screen.

<u>Information provided to participant before variable ART teammate practice and official rounds:</u>

You will now be working with your variable response teammate. It will respond based on the rate at which new ships appear. In other words, it will respond more rapidly as ships appear at a faster rate, and less rapidly as ships appear at a slower rate.

References

Bindewald, J. M. (2015). *Adaptive Automation Design and Implementation*. *Theses and Dissertations*. Retrieved from https://scholar.afit.edu/etd/211

Bindewald, J. M., Miller, M. E., & Peterson, G. L. (2014). A function-to-task process model for adaptive automation system design. *International Journal of Human-Computer Studies*, *72*(12), 822–834. https://doi.org/10.1016/j.ijhcs.2014.07.004

Bindewald, J. M., Miller, M. E., & Peterson, G. L. (2019). Creating Effective Automation to Maintain Explicit User Engagement. *International Journal of Human-Computer Studies*.

Department of Defense: Defense Science Board. (2012). Task Force Report: The role of autonomy in DoD Systems. Washington, D.C.: Office of the Under Secretary of Defense for Acquisition, Technology and Logistics.

Endsley, M. R. (2015). *Autonomous Horizons: System Autonomy in the Air Force - A Path to the Future (Vol 1)*. Retrieved from https://www.af.mil/Portals/1/documents/SECAF/AutonomousHorizons.pdf

Fahey, K. M., & Miller, M. J. (2017). Unmanned Systems Integrated Roadmap 2017-2042. Department of Defense.

Gettinger, D. (2017). *Drones in the Defense Budget: Navigating the Fiscal Year 2018 Budget Request*. Retrieved from http://interactive.africandefence.net/dronesinthedrc/?mc_cid=5a72d197e1&mc_eid=b57909e368

Goodman, T. J. (2016). *Understanding Effects of Autonomous Agent Timing on Human-Agent Teams Using Iterative Modeling, Simulation and Human-in-the-Loop Experimentation*. Air Force Institute of Technology. Retrieved from http://www.dtic.mil/docs/citations/AD1054087

Goodman, T. J., Miller, M. E., Rusnock, C. F., & Bindewald, J. M. (2017). Effects of agent timing on the human-agent team. *Cognitive Systems Research*, *46*, 40–51. https://doi.org/10.1016/j.cogsys.2017.02.007

Kearns, K. (2015). RFI: Autonomy for Loyal Wingman. Air Force Research Laboratory (AFRL).

Parasuraman, R. (2008). *Supporting Battle Management Command and Control: Designing Innovative Interfaces and Selecting Skilled Operators*. Fairfax, VA. Retrieved from http://www.dtic.mil/docs/citations/ADA480645

Raibert, M., Blankespoor, K., Nelson, G., & Playter, R. (2008). Bigdog, the Rough-Terrain Quadruped Robot. In *unair.ac.id*. Seoul, Korea. Retrieved from http://web.unair.ac.id/admin/file/f_7773_bigdog.pdf

Schneider, M. F., Bragg, I. L., Henderson, J. P., & Miller, M. E. (2018). Human Engagement with Event Rate Driven Adaptation of Automated Agents. In *2018 IISE Annual Conference*. Orlando, FL.

Sterling, L., & Taveter, K. (2009). *The art of agent-oriented modeling*. Cambridge, MA: The MIT Press.

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
21-03-2018	Master's Thesis	August 2017-March 2019

4. TITLE AND SUBTITLE	
Assessing Artificial Agent Response Time Effects on Human-Agent Teams in Variable Inter-Arrival Time Environments	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER** 2016-067
Canzonetta, David J., First Lieutenant, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865	AFIT-ENV-MS-19-M-166
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFOSR
US Air Force Office of Scientific Research Attn: Dr. James Lawton 4075 Wilson Blvd., Suite 350 Arlington, VA 22203 (703) 696-5999 (DSN: 426-5999), james.lawton.1@us.af.mil	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Autonomous systems have gained an expanded presence within the Department of Defense (DoD). Furthermore, the DoD has clearly stated autonomous systems must extend the capabilities of their human operators. Thus, the exploration of strategies for effective pairing of humans and automation supports this vision. Previous research demonstrated that the time at which an automated agent assumes a task for its human teammate, or agent response time (ART), affects human-agent team performance, human engagement, and human workload. However, in this research environment, the time between subsequent tasks appearing to the human-agent team, or inter-arrival time (IAT), remained constant. Variable IAT environments more accurately reflect real-world operational environments. Previous research also maintained ART at a fixed level. Additionally, the effect of human understanding of automated teammate actions on human-agent team performance remains unknown.

This thesis attempts to analyze the effect of an agent with adaptive ART that varies based on current IAT on human-agent team performance, human engagement, and human workload. Additionally, it seeks to determine the implication of agent predictability to the human. This thesis explores these issues in three phases. First, a method and development of a variable ART function for use in future phases is presented. Second, a study of a variable ART teammate against a fixed ART teammate highlights the significance of providing detailed agent instruction to the human. Third, analysis of instruction and type of agent teammate across an entire input IAT function and at different IAT levels is conducted. This work establishes key factors for adaptive ART function implementation. Based on specific IAT changes, the current research demonstrates that adaptive ART can boost human-agent team performance and manipulate human engagement. Furthermore, predictability of agent action in variable IAT environments is a desired system attribute.

15. SUBJECT TERMS

Human-Agent Teaming; Adaptive Agent Response Time; Agent Predictability

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Michael E. Miller, PhD, AFIT/ENV
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	62	(937) 255-3636, ext 4561 (Michael.Miller@afit.edu)