Air Force Institute of Technology
AFIT Scholar

Theses and Dissertations Student Graduate Works

3-2020

Human Performance Modeling: Analysis of the Effects of Manned-Unmanned Teaming on Pilot Workload and Mission Performance

Jinan M. Andrews

Follow this and additional works at: https://scholar.afit.edu/etd

Part of the Systems Engineering Commons

Recommended Citation
Andrews, Jinan M., "Human Performance Modeling: Analysis of the Effects of Manned-Unmanned Teaming on Pilot Workload and Mission Performance" (2020). *Theses and Dissertations*. 3225.
https://scholar.afit.edu/etd/3225

This Thesis is brought to you for free and open access by the Student Graduate Works at AFIT Scholar. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of AFIT Scholar. For more information, please contact AFIT.ENWL.Repository@us.af.mil.

HUMAN PERFORMANCE MODELING: ANALYSIS OF THE EFFECTS OF MANNED-UNMANNED TEAMING ON PILOT WORKLOAD AND MISSION PERFORMANCE

THESIS
MARCH 2020

Jinan M. Andrews
2nd Lieutenant, USAF

AFIT-ENV-MS-20-M-185

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

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

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

AFIT-ENV-MS-20-M-185

HUMAN PERFORMANCE MODELING: ANALYSIS OF THE EFFECTS OF MANNED-UNMANNED TEAMING ON PILOT WORKLOAD AND MISSION PERFORMANCE

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Systems Engineering

Jinan M. Andrews, MS

2nd Lieutenant, USAF

March 2020

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

AFIT-ENV-MS-20-M-185

HUMAN PERFORMANCE MODELING: ANALYSIS OF THE EFFECTS OF MANNED-UNMANNED TEAMING ON PILOT WORKLOAD AND MISSION PERFORMANCE

Jinan M. Andrews, MS

2nd Lieutenant, USAF

Committee Membership:

Lt Col Christina F. Rusnock, PhD
Chair

Michael E. Miller, PhD
Member

Douglas P. Meador, PhD
Member

AFIT-ENV-MS-20-M-185

Abstract

Due to the advent of autonomous technology coupled with the extreme expense of manned aircraft, the Department of Defense (DoD) has increased interest in developing affordable, expendable Unmanned Aerial Vehicles (UAVs) to become autonomous wingmen for jet fighters in mosaic warfare. Like a mosaic that forms a whole picture out of smaller pieces, battlefield commanders can utilize disaggregated capabilities, such as Manned-Unmanned Teaming (MUM-T), to operate in contested environments. With a single pilot controlling both the UAVs and manned aircraft, it may be challenging for pilots to manage all systems should the system design not be conducive to a steady state level of workload.

To understand the potential effects of MUM-T on the pilot's cognitive workload, an Improved Performance Research Integration Tool (IMPRINT) Pro pilot workload model was developed. The model predicts the cognitive workload of the pilot in a simulated environment when interacting with both the cockpit and multiple UAVs to provide insight into the effect of Human-Agent Interactions (HAI) and increasing autonomous control abstraction on the pilot's cognitive workload and mission performance. This research concluded that peaks in workload occur for the pilot during periods of high communications load and this communication may be degraded or delayed during air-to-air engagements. Nonetheless, autonomous control of the UAVs through a combination of Vector Steering, Pilot Directed Engagements, and Tactical Battle Management would enable pilots to successfully command up to 3 UAVs as well as their own aircraft against 4 enemy targets, while maintaining acceptable pilot cognitive workload in an air-to-air mission scenario.

Acknowledgments

First of all, I would like to extend my sincere appreciation to my advisor, Lt Col Christina Rusnock, and co-advisor, Dr. Michael Miller, for their selfless guidance and support over the course of my time at AFIT. I also thank my committee member, Dr. Douglas Meador, for sharing his expertise and providing direction. I would also like to thank Captain Tyler Goodman and the 711 Human Performance Wing at Wright-Patterson AFB; this research would not be possible without their critical research, cooperation, and guidance. Finally, I would like to thank my family and great friends for their nonstop love and support throughout this process.

Jinan M. Andrews, 2nd Lieutenant, USAF

Table of Contents

List of Figures

List of Tables

HUMAN PERFORMANCE MODELING: ANALYSIS OF THE EFFECTS OF MANNED-UNMANNED TEAMING ON PILOT WORKLOAD AND MISSION PERFORMANCE

I. Introduction

Chapter Overview

This chapter begins by covering the background of Unmanned Aerial Vehicles (UAVs) and introducing the topic of autonomous wingmen for jet fighters. It then focuses on the effect of Human-Agent Interactions (HAI) and autonomous control on the pilot's cognitive workload during flight operations. Next, the chapter explains how Improved Performance Research Integration Tool (IMPRINT) can help predict pilot workload and mission performance when interacting with both the ownership cockpit and the UAVs. After the research and investigative questions have been presented, this chapter then focuses on the best course of action to address the research problem. Lastly, the chapter addresses the assumptions and limitations, research implications, and provides a preview of the remaining chapters.

Introduction of Manned-Unmanned Teaming in Air Warfare

The rise of adversaries in combat air space has motivated the United States military to actively explore experimental flight alternatives to attempt to augment America's fighter squadrons. With the foreseeable future for air warfare leaning towards the use of UAVs, the Department of Defense (DoD) is investigating the use of UAVs to augment manned tactical platforms, with the goal of enhancing capabilities for operating in or permissive through contested airspace. To accomplish this, a UAV concept dubbed Manned-Unmanned Teaming (MUM-T) is being explored where the UAVs will act as teammates to human pilots in air operations and address current operational limitations and perhaps improve human survivability in modern warfare (Drew, 2016). The lower cost of the UAVs, as compared to manned aircraft, has received increased attention by the U.S. military due to its potential to expand the combat capacity of manned fighters and bombers within the limitations of the DoD's budget. Experimental technologies such as the Air Force Research Laboratory's (AFRL) XQ-58A Valkyrie, Boeing's Airpower Teaming Systems (ATS), and Kratos Defense & Security Solutions' Unmanned Tactical Aerial Platform-22 (UTAP 22) could potentially provide autonomous jet fighters for a fraction of the price of a F-35 Lightning II Joint Strike Fighter or F-22 Raptor aircraft (Hanlon, 2017). The emergence of this technology presents a low-cost solution that shifts the paradigm of a pilot commanding a single aircraft to a pilot commanding multiple UAVs in addition to the manned aircraft. Using MUM-T in air operations would alter the warfighter Concept of Operations (CONOPS) and traditional life cycle management paradigms.

In theory, a manned aircraft would be paired with one or multiple robotic wingmen to act in unison with the command pilot to locate, jam, strike, or distract enemy air defenses (Rogoway, 2017). The UAVs would operate at a far off distance to provide pilots with additional weapons and sensors while increasing the enemy's targeting requirement in the battlefield. The unmanned aircraft could carry out surveillance missions and amplify firing power to fill capacity gaps for pilots. It also enables airmen to access new areas of the battlespace that may be too difficult or risky for a human pilot to enter. These additional capabilities make MUM-T a potentially lethal force and a significant asset to the military. The DoD recognizes these potential advantages and has taken steps towards exploring the potential of these affordable, unmanned tactical aircraft.

However, there are complications with this new strategy, should the DoD choose to adopt MUM-T for frontline use. The command pilot bears the weight of the combat effort and will need to deploy capabilities from the UAV in addition to commanding their own aircraft, ideally without degrading the effectiveness of their own aircraft within the mission. This concept places additional cognitive demands on the pilots, potentially exceeding their available resources should the system interface design not be conducive to maintaining a manageable level of workload for the pilot. The challenge of maintaining close and time critical control of UAVs requires a new approach to control and integration. The DoD must re-evaluate some of the basic conventions in current operations to leverage the best of traditional aviation and emerging capabilities. By further investigating the effects of HAI and autonomous control on the pilot's cognitive workload and mission performance, this study seeks to provide insight into the impact of MUM-T on the command pilot in air-to-air operations.

Problem Statement

The level of success achieved through MUM-T is highly dependent on the integration of this technology with human operators. Researchers have studied the design of autonomous systems within remote controlled flight. However, there is limited research investigating workload impacts of more autonomous technology in military flight operations. This is likely due to the novelty of MUM-T. These systems will require an improved understanding of operator mental workload and how it affects mission performance to enable successful integration of pilots and UAVs into a single cohesive, effective team.

To support informed decisions on the available operations concepts associated with MUM-T, a thorough and in-depth study of the effects of MUM-T on the pilot's cognitive workload and mission performance is required. This is a significant area to explore because the structure of the human-agent system affects the human's cognitive workload and thus, the human-agent team's overall effectiveness in combat. The simulation developed as part of this analysis was designed to provide a method to evaluate the effects of HAI and autonomous control on the pilot's cognitive workload and mission performance. This research seeks to identify workload management strategies and a preferred design for the control and integration of UAV technologies in manned operations.

Research Objectives

The purpose of the thesis was to understand the potential effects of MUM-T on the pilot's cognitive workload and overall mission performance. There were two main objectives to this study:

1. The first objective was to develop an original Discrete Event Simulation (DES) within IMPRINT that quantitatively models the mental workload of pilots during flight operations with UAVs to reveal any potential benefits or issues from the HAI.
2. The second objective was to determine what amount of autonomous control abstraction has the largest impact in reducing operator workload and increasing system performance to provide HAI recommendations for system improvements.

Investigative Questions

The following research questions were addressed to fully answer the overarching inquiry of how to model the MUM-T system such that the HAI and can be investigated to study its potential effects on pilot workload and mission performance:

1. ***How does the use of MUM-T affect the pilot's cognitive workload during combat mission events?***

   The first question was used to determine if the relationship between the deployment of UAVs and workload metrics are linear or non-linear. It was hypothesized that the deployment of UAVs in air operations would result in higher workload than situations where the pilots did not need to command the UAVs and their aircraft.

2. ***How does the use of MUM-T affect the human-agent team's mission performance during combat mission events?***

   The second question was used to determine how the incorporation of UAVs into air operations would impact the human-agent team's overall mission performance in terms of enemy target kills. It was hypothesized that the utilization of UAVs in air operations would improve the human-agent team's ability to successfully strike targets.

3. ***To what degree of autonomous control abstraction should the UAVs perform at to reduce operator workload in a flight operation task?***

   The third question was used to determine how much of the operator's cognitive tasks should be relinquished by the command pilot and reassigned to the UAVs to reduce the amount of workload experienced by the pilot. It was hypothesized that the pilot's workload levels would be reduced by commanding the UAVs to meet a desired goal and enabling the MUM-T system to make all required decision to meet those goals through Tactical Battle Management.

4. ***To what degree of autonomous control abstraction should the UAVs perform at to increase mission performance in a flight operation task?***

   The fourth question was used to determine how much of the operator's cognitive tasks should be relinquished by the command pilot and reassigned to the UAVs to help and not hinder the human-agent team's mission performance. It was hypothesized that the human-agent team's mission performance would also be improved by commanding the

UAVs to meet a desired goal and enabling the MUM-T system to make all required decision to meet those goals through Tactical Battle Management.

Methodology

To explore the decision to integrate an automated component into a human system, this study built an original DES using IMPRINT to research the effect of MUM-T on the pilot's cognitive workload and mission performance. IMPRINT is a discrete event modeling tool specifically designed to evaluate the interactions of human users and system technologies (Rusnock & Geiger, 2013). It was developed by ALION and funded by the U.S. Army Research Laboratory, Human Research & Engineering Directorate, to support manpower and personnel integration as well as human systems integration (Alion Science and Technology Corporation, 2009). The tool models human workload and performance as a function of time by tracking activities performed by the human or machine. It can test multiple alternate scenarios in a short period of time as well as quantify the effect of a system interface design on the human element of a system based on mental workload. This type of evaluation is useful for gauging the effect of HAI and autonomous control on the pilot's cognitive workload and mission performance.

Although IMPRINT is not yet widely used for human-agent systems, it is possible to model existing operation procedures and inputs from external stimuli (i.e. UAVs flying around a fighter aircraft) to observe and predict workload levels through computer simulation. This study developed a DES that was constructed from data gathered from Autonomy for Air Combat Missions (ATACM), a separate study previously performed by the 711th Human Performance Wing (HPW) at AFRL, Wright Patterson Air Force Base. The ATACM study was a Human-In-the-Loop (HITL) experiment that developed and tested critical autonomous decision and machine learning technologies in a virtual simulation cockpit with the aim of enabling a single pilot to command multiple UAVs in flight while controlling his or her own aircraft in highly contested environments (Schumacher et al., 2017). The study replicated an offensive counter-air scenario in which individual performance and mental workload could vary in real-time based on the operators' capabilities.

Using the ATACM study, an original DES was constructed to model the mission scenarios and system configuration assumed within this assessment. A baseline DES was developed to quantitatively capture the pilot's cognitive workload levels and mission performance when controlling both UAVs and manned aircrafts. Alternative system configurations were then created to compare the baseline model to varying amounts of autonomous control abstraction and traditional aviation techniques. Through this process, this research sought to understand and determine how integrating UAVs into flight operations impacts the command pilot's workload and mission performance. The findings presented in this research are a significant step towards simulating the complexities of real-world activities by mirroring the highly dynamic nature of realistic military operations in a virtual environment.

Assumptions and Limitations

Creating an IMPRINT model required task analyses, direct observations, and data collection of a system. However, MUM-T has yet to be deployed in an operational environment. Consequently, this research was reliant on information provided by Subject Matter Experts (SMEs) and data collected from a HITL study performed by the 711th HPW. An in depth analysis of the assumptions and limitations of this research is provided in the final discussion chapter.

Research Implications

This research is expected to have a significant impact on projects, such as AFRL's Skyborg and Autonomous Collaborative Platforms programs, which are currently developing integrated, human-agent aircraft systems for operational use. The results of this study delivered a cost-effective way to evaluate MUM-T systems without having to perform costly, time-consuming HITL experiments. Furthermore, the study provided valuable insight into the effects of incorporating UAVs into air operations, which can then be used to refine UAV requirements before fielding the unmanned combat air vehicle. This research ultimately has the potential to refine the relationship between pilots and UAVs to lead to a more nuanced understanding of how to best incorporate MUM-T into military air warfare.

Preview

This research follows the scholarly format, thus some of the chapters are self-contained drafts of potential publications. This chapter began with the background of MUM-T and described a problem that needs to be addressed within human-agent teaming. Chapter II contains a literature review from relevant sources on the topics of automation, mental workload, DES in aviation. Chapter III addresses the first research objective by investigating the effects of HAI on the pilot's cognitive workload and mission performance when incorporating UAVs in an air-to-air operation. Chapter IV addresses the second research objective by identifying the stages and levels of automation that have the largest impact in reducing operator workload and increasing system performance. Chapter V contains a summary of the research results and future research recommendations.

II. Literature Review

Chapter Overview

The purpose of this chapter is to provide relevant background information from previous research and important works of literature to foster an understanding of the topics discussed in this research. The chapter begins by providing a generalized overview of automation to include its advantages and disadvantages as well as the stages and levels of autonomous control. It then describes the effect of autonomous control on system design and performance. The chapter subsequently dives into workload theory by explaining what it is, how it relates to human performance, and how it can be measured. Finally, the researcher introduces IMPRINT, which is useful in quantitatively modeling the mental workload of operators. The chapter concludes by stating the research gap that this work fulfills and closes with a short conclusion on all of the topics that were discussed. Each subject is described in detail to establish the intellectual foundation of the subject areas necessary to follow the discussion throughout the thesis chapters.

Automation

Autonomous control and automation go hand-in-hand, boosting and providing a fallback for one another. Autonomous control is the self-governance of control functions amidst significant uncertainties in the environment and the ability to compensate for system failures without external intervention (Antsaklis, Passino, & Wang, 1991). This is different from automation, which is often defined as a process or procedure performed with minimal human assistance (Groover, 2015). Automation is also defined as the capability of a machine or computer agent (hereafter referred to as “agent”) to execute a task previously performed by a human operator (Parasuraman, Sheridan, & Wickens, 2000). Examples might include a calculation performed by a computer instead of a human or the ability for a machine to make decisions without human intervention. The degree of complexity can vary in automation, ranging from organizing information sources, to recommending options, or perhaps carrying out an action. In each of these cases, automation serves to fulfill the functions of the human operator at varying levels of control.

Automation has played a key role in the technological development of modern day aircraft systems. Advancements in computer software and hardware have enabled aviation systems to perform simple to complicated tasks that human operators performed in the early days of aviation. To understand the evolution of flight management systems, it is important to recognize the fundamentals of automation, to include what it is, the advantages and disadvantages of automation, as well as the models and levels of autonomous control.

Advantages and Disadvantages of Automation

Human-agent teaming is the cooperation between one or more people and intelligent agents, capable of dynamically engaging with one another for the purpose of achieving a common goal that is beneficial to the mission. The concept of "intelligent agents" implies the independent ability to sense, reason, and act upon the environment. Thus, inferring that intelligent agents have a higher level of adaptability and flexibility than non-intelligent agents, enabling them to vary their performance in response to environmental factors.

The MUM-T concept is an example of human-agent teaming. The UAVs will act like assistants to human pilots in air operations by bolstering defense networks and aiding in certain classes of decision making. This capability, as with other automated systems, can provide several advantages and disadvantages to the human operator. In general, automation can reduce human task load or increase operator efficiency by relieving the operator from specific tasks. For instance, the agent could perform complex mathematical calculations, organize or filter information for relevance and coherency, perform mundane or routine tasks, or monitor a system for an extended amount of time, thus reducing human participation, information overload, and consequently human error (Parasuraman et al., 2000; Swanson et al., 2012). These benefits are ideally obtained when a balance is struck between the capabilities of the system, what the system can achieve, and the demands placed on the human resources (Taylor, 2006). In these situations, automation not only improves safety by reducing human error, but also increases reliability, improves precision, and reduces operator workload (Billings, 1991; Hart & Sheridan, 1984). Furthermore, operator fatigue accumulates more slowly and the human operator has a greater capacity to perform more critical tasks as a result of reducing operator workload (Secarea, 1990). For these reasons, automation that is well-designed can amplify operator's capabilities in the cockpit as well as in other human-agent teaming systems.

Despite these advantages, not all systems that can be automated should be automated (Wiener & Curry, 1980). Automation can help reduce issues such as human error or information overload, but clumsy automation can also create several new problems such as operator complacency, boredom, decision-bias, trust issues, as well as increase fluctuations in workload (De Visser et al., 2008; Woods, Johannesen, Cook, & Sarter, 1994). First of all, automation may cause an operator to become complacent because the operator's interaction with the system is reduced to a monitoring role. This change can lead to a loss of manual skills, system knowledge, and even job satisfaction (Hart & Sheridan, 1984; R. D. Johnson, Bershader, & Leifer, 1983). The operator's situation awareness is degraded when automation takes over all processes, especially when the information applied by the operator is not readily available to the operator. Secondly, the lack of appropriate communication in poorly designed automation can lead to operator distrust or confusion (Endsley, 1996). If the human is missing vital pieces of information about the process or system state (i.e. automation's logic, functionality, responsibilities, limits, state, or operating parameters), then it will be difficult for the person to understand what the system is doing or why it is doing it (Wiener, 1989). The breakdown of communication between the operator and automation may lead to decision bias and/or trust issues between the operator and the automated system.

Another disadvantage of automation arises when new burdens are inadvertently placed on the operator. Automation can eliminate human tasks in some circumstances, but also generate new tasks or problems in conjunction with the expected benefits of automation; consequently, adding more opportunities for error or increasing operator workload (Colombi et al., 2011; Woods et al., 1994). For instance, automation could increase workload because of the added communication between the system and the operator or the replacement of physical control activities with supervisory activities (Endsley, 1996). Moreover, automation can contribute to hazardous attitudes such as misuse (using automation when it should not be used), disuse (not using available and capable automation), or abuse (inappropriate use of automation) (Parasuraman & Riley, 1997). It is important to understand the disadvantages of automation because all of these issues add another dimension of complexity to the design of human-agent systems.

Stages and Levels of Automation

To effectively leverage the advantages of automation, designers should be aware of the varying degrees of autonomous control in human-agent teams. Automation can operate across a spectrum of autonomous control defined in Table 1-Table 3 and Figure 1-Figure 2. Although these hierarchies focus on what to automate and how to allocate functions, there are interdependencies between humans and agents. For this reason, the stages and levels of automation to be covered in this section are flexible and can be synthesized to take into account cross-scale interactions. The four types of autonomous control taxonomies are listed below.

1. Ten Levels of Automation (Sheridan & Verplank, 1978)
2. Four Stages of Human Information Processing (Parasuraman et al., 2000)
3. Five Levels of Decision Automation (Hart & Sheridan, 1984)
4. Four Levels of Allocation of Roles Between the Expert System and The Pilot (Endsley, 1987)

In 1978, Sheridan and Verplank described the distribution of tasks allocated between either the human or the automation in the ten LoA. This ten-point scale characterizes the level of involvement granted to automation within human-machine or human-agent teams by using a continuum of levels, ranging from no automation (i.e. human manually performs task) to full automation (i.e. computer is fully autonomous). Table 1 describes the ten LoA where higher levels represent increasing automation autonomy over human actions (Parasuraman et al., 2000; Wickens, Mavor, & McGee, 1998).

Table 1. Ten Levels of Automation – adapted from (Sheridan & Verplank, 1978)

	Level	**Description**
Low	1	Fully manual control; computer offers no assistance; human does all planning, decision making, and action execution
	2	Computer provides a complete set of decision/action alternatives, or
	3	Narrows the selection down to a few, or
	4	Suggests one alternative, and
	5	Executes that suggestion if the human approves, or
	6	Allows human limited time to veto decision before automatic execution
	7	Executes automatically, then necessarily informs the human, and
	8	Informs the human upon request, or
	9	Informs the human only if it, the computer, decides to
High	10	Fully autonomous control; computer decides everything and ignores the human

The ten LoA range from complete human control to complete computer control. The amount of decision authority granted to the automation increases as the level of the scale increases. At level 1, there is no automation because the operator executes all of the tasks. At level 4, the computer suggests one decision alternative from the provided options, but the human has the final decision authority. At level 6, the human is only given a limited amount of time to veto a decision before the computer carries out its decision. At level 10, the system is fully automated and there is no human interaction because the computer has full control to make and execute a decision. As the levels increase, the amount of approval authority required before an artificial agent initiates an action decreases. Consequently, Sheridan's and Verplank's LoA illustrates how operator involvement decreases as automation is granted the authority to perform tasks traditionally performed by humans (Vagia, Transeth, & Fjerdingen, 2016).

To understand the different ways automation can be applied to a system, Parasuraman, Sheridan, and Wickens used the four-stage model of Human Information Processing (HIP) to re-examine tasks at a detailed level (Broadbent, 1958; Parasuraman et al., 2000). The HIP model is composed on four stages: 1) sensory processing; 2) perception/working memory; 3) decision making; and 4) response selection (Parasuraman et al., 2000). The four stage model is shown in Figure 1.

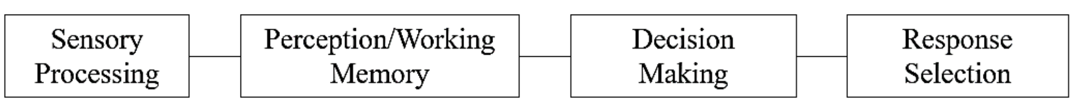


Figure 1. Human Information Processing Model– adapted from (Parasuraman et al., 2000)

As automation replaces human operated tasks, the replaced tasks may relate to any of the four stages of the HIP model. Parasuraman et al. introduced the idea of associating LoA to the HIP by translating the stages into four corresponding system functions: 1) information acquisition; 2) information analysis; 3) decision and action selection; and 4) action implementation. When the stages are assigned to a system, the resulting functions provide an initial categorization for the types of tasks in which automation can support the human operator. The relationship between the two processing models is shown in Figure 2.

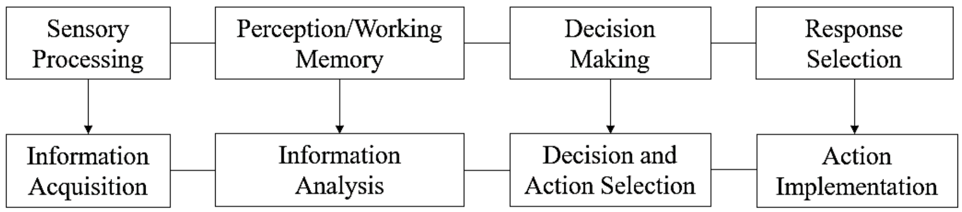


Figure 2. Stages of Machine Processing Built from the Human Information Processing Model – adapted from (Parasuraman et al., 2000)

The four stages of HIP describe human decision-making and the functions correlate with system processing. In the first stage, sensory processing, information is gathered from the outside world and used for higher level processing. Information acquisition supports sensory processing by controlling sensors and the registration of multiple sources of input data. This step includes the orienting of sensory receptors, sensory processing, selective attention, and initial pre-processing of data prior to full perception (Parasuraman et al., 2000). In the second stage, perception/working memory, information that is gathered from the first stage is synthesized in consort with long-term memory to form an interpretation of the environment. Information analysis supports working memory and inferential processes by conscious perception, filtering the retrieved raw data, and processing it into information that is more important or useful for the human (Baddeley, 1996). This step includes cognitive operations such as rehearsal, integration and prediction, but these operations occur prior to the point of a decision (Parasuraman et al., 2000). In the third stage, decision making, a course of action is selected from the different decision alternatives based upon the interpretation of the environment. The decision and action selection supplements cognitive processing and human decision abilities by presenting a desired choice to the human without taking that action (Parasuraman et al., 2000). In the final stage, response selection, the response or action decided upon in the decision making stage is executed (Kaber, Stoll, & Thurow, 2007; Parasuraman et al., 2000). It is the actual execution of the action choice. By and large, the LoA across any of the four stages and functions of automation can vary in design and application, depending on the demands and uses of the operational system according to the proposed model.

Since Sheridan and Verplank, several researchers have proposed alternate taxonomies describing LoA (Clough, 2002; Draper, 1995; Endsley, 1987; Endsley & Kaber, 1999b; Endsley & Kiris, 1995; Fereidunian, Lehtonen, Lesani, Lucas, & Nordman, 2007; Fereidunian, Lucas, Lesani, Lehtonen, & Nordman, 2007; Hart & Sheridan, 1984; M. Johnson, Bradshaw, & Feltovich, 2018; Kaber, 2018; Lorenz et al., 2001; Milgram, Rastogi, & Grodski, 1995; Ntuen & Park, 1988; Proud, Hart, & Mrozinski, 2003; Riley, 1989). Each of these authors proposed varying LoA for different taxonomies. What is important to remember is that even taxonomies that are supposed to be used for the same types of applications can vary a lot (Vagia et al., 2016). For example, automation allocation for avionics can be explained by Hart and Sheridan's (1984) five Levels of Decision Automation, shown in Table 2, or Endsley's (1987) four levels of Allocation of Roles, shown in Table 3.

Table 2. Five Levels of Decision Automation – adapted from (Hart & Sheridan, 1984)

	Level	**Description**
Low	1	Automated system suggests alternatives for human to consider or ignore
	2	Automated system lists alternatives from which human must decide and execute manually
	3	Automated system lists alternatives from which human must decide, but system executes
	4	Automated system makes decision, but informs human who can intervene before execution of decision
High	5	Automated system makes decision and executes, only informing human after the fact

Table 3. Four Levels of Allocation of Roles – adapted from (Endsley, 1987)

	Level		**Description**
Low	1	Suggest	Pilot chooses whether or not to act upon expert system recommendations
	2	Concur	Expert system acts autonomously, however, the consent of the pilot is required to carry out actions
	3	Veto	Expert system act autonomously, unless recommendations are vetoed by the pilot
High	4	Act	Fully autonomous with no operation interaction; expert system excludes pilot from the loop

Both taxonomies are supposed to be used for the same application, avionics decision support, however these scales differ in the number of levels their taxonomies include. The model presented by Hart and Sheridan (1984) describes the LoA in five levels ranging from autonomous suggestions to fully autonomous control, with the exception of a fully manual control level and fewer intermediate levels. While the compact model proposed by Endsley (1987) presents four functions for the allocation of roles between an advanced cockpit (i.e. expert system), capable of supplementing human decision making, and the operator (i.e. pilot). In this sense, the designer has the freedom to decide which LoA approach fits best to his or her needs as there is not one prescribed way to design an autonomous system.

Due to the growth of automation capabilities in recent years, a new set of human-agent design tools have been proposed to keep up with the advancement of sociotechnical systems (Allen, Guinn, & Horvitz, 1999; Fong, Thorpe, & Baur, 1999; C. D. Johnson, Miller, Rusnock, & Jacques, 2017; M. Johnson, Bradshaw, et al., 2018; M. Johnson et al., 2014a; M. Johnson, Vignati, & Duran, 2018; Klein, Woods, Bradshaw, Hoffman, & Feltovich, 2004; Miller, 2017). Miller (2017) discussed the problem of automatically coordinating the behaviors of multiple agents to achieve more complex goals, since such a feat would require high precision coordination – a difficult task to achieve. Instead, he recommended that human-agent interaction should strive to adopt a more explicit interaction protocol to help coordinate roles and responsibilities. According to this view, this would make collaborative task performance feasible in complex domains.

Johnson, Bradshaw, et al. (2018) made a different observation that traditional approaches often drive designers towards deciding what to automate as if it were a binary decision. However, they made the point that these two cases are “degenerate cases where the situation does not permit coordination” (Johnson, Bradshaw, et al., 2018). According to this view, LoA-based approaches could be characterized as restrictive, forcing designers to choose what to automate and how to allocate functions, instead of leading them to coordinate the task work in support of interdependencies between humans and automation.

In response to these concerns, Johnson et al. (2017) developed the five Levels of Human Control Abstraction (LHCA) as an alternative conceptual framework to describe the level of control inputs given by the operator (see Table 4). The framework describes the cognitive tasks relinquished by the human operator and reassigned to the automation. As the stages progress from Direct Control to Mission-Capable Control, the level of detail for the human operator's control inputs, attention, and workload is reduced. For instance, an example of Direct Control would be a simple, fixed wing aircraft, whereas an example of Mission-Capable Control would be an autonomous car. Using this taxonomy, designers have a better understanding of the workload that is placed on the human operator when interacting with the automation in addition to the levels of human control abstraction for each interaction. However, a weakness of this model is that there is not enough precision to fully capture the nuances between each LHCA.

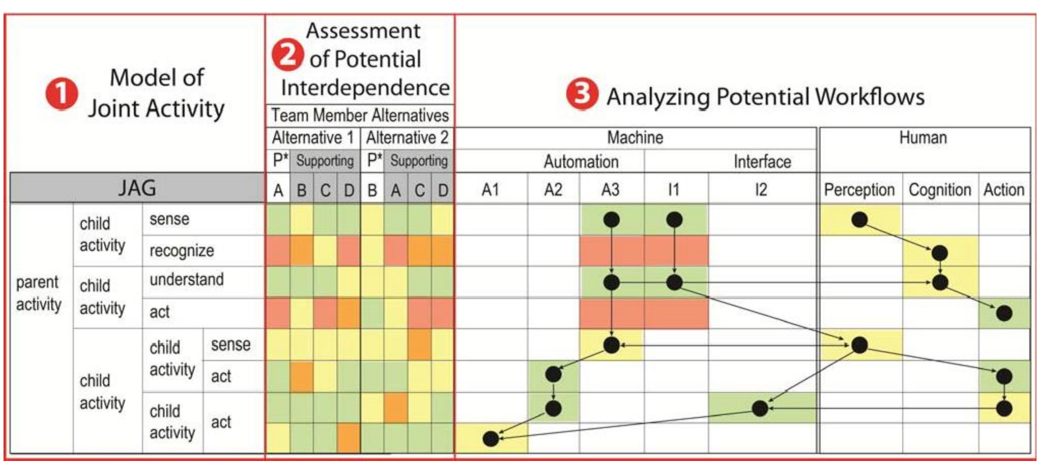
Table 4. Levels of Human Control Abstraction – adapted from (C. D. Johnson et al., 2017)

	Level		**Description**
Low	1	Direct Control	Operator controls every aspect of the system, including actual control surface positions or motor power
	2	Augmented Control	Operator gives control inputs commanding desired actions, the system then makes final determinations about control surface positions or motor power
	3	Parametric Control	Operator inputs desired parameters that the system should meet, the system then uses onboard sensors and control algorithms to meet those parameters
	4	Goal-Oriented Control	Operator inputs desired goals the system should meet, the system then makes all required decisions to meet those goals
High	5	Mission-Capable Control	Operator enters pre-launch mission goals at a level of detail which, when combined with standard operating procedures and rules of engagement, are sufficient to accomplish the mission

The five LHCA can be modelled using the Interdependence Analysis Tool (IAT) (Johnson, Vignati, et al., 2018). The IAT is like a road map that helps designers visually understand how people and automation can effectively team by providing insight into the interdependence relationships used to support one another throughout an activity. The IAT allows designers to track which entity in the human-machine system is performing each specific sub-task across multiple activities and how the workflow changes over time. This is beneficial because it helps designers to see the changes in role assignment between the human and the machine. However, unlike the LHCA model, IAT does not map the level of workload placed on the human or the machine in each activity.

The IAT was founded upon three essential interdependence relations: observability, predictability, and directability (Johnson et al., 2014). From this foundation, Johnson et al. (2018) developed an experimental paradigm containing three main sections: 1) joint activity modelling, 2) assessment of potential interdependence, and 3) analysis of potential workflows. Table 5 illustrates these three main sections in a generic table. Section 1 helps designers model the joint activity, section 2 helps them identify potential interdependencies in the activity, and section 3 helps analyze the potential workflows to better understand the flexibility and risk in the human-machine system (Johnson, Vignati, et al., 2018). Ideally, the amount of mental workload experienced by the operator decreases as the responsibilities shift from the human to the machine. This shift in responsibilities from the human to the machine can be seen in third section of the IAT.

Table 5. Interdependence Analysis Tool – from (M. Johnson, Vignati, et al., 2018)



In review, there have been several taxonomies proposed for the stages and levels of autonomies over the past four decades. Each model has its unique nuances, but all work toward the goal of providing a language to characterize the division of work between the human and the agent. However, their application often leads the designer to select a fixed allocation, which may not always be appropriate as illustrated by the writings of Johnson et al. (2018) and C.D. Johnson et al. (2017). System designers need to be able to evaluate what type of interaction or interdependence between the human and the automation is most appropriate for a human-agent team as well as identify when automation should be utilized to maximize the use of its capabilities.

Effects of Levels of Automation

While automation may lead to legitimate system advantages, quantification of these advantages should include the whole system including the operator's cognitive workload, situation awareness and the effect of these attributes on mission performance. Several studies have been conducted to explore the effects of LoA on human workload, situation awareness, and system performance within real world or simulated systems (Cummings, Bruni, Mercier, & Mitchell, 2007; Endsley & Kaber, 1999; Kaber & Endsley, 2004; Mitchell, 2000; Parasuraman et al., 2000). Kaber and Endsley's research in 1999 and 2004 found that LoA is an important factor in determining the overall performance of a human-agent system. According to their studies, workload remains stable, situation awareness is degraded, and overall system performance improves as the LoA is increased from low to intermediate (Endsley & Kaber, 1999; Kaber & Endsley, 2004). It was determined through this research that if the designer automated higher level cognitive functions, the operator may experience underload and lose focus on task execution. Consequently, decreasing the operator's situation awareness and negatively impacting the human-agent team's performance. Conversely, if the designer only incorporated lower LoA, then the operator's cognitive workload could become excessive and negatively impact overall system performance. Kaber and Endsley's results illustrate the complex relationship between cognitive workload, situation awareness, and system performance (Endsley & Kaber, 1999; Kaber & Endsley, 2004).

It is essential that researchers understand the potential effects of LoA when designing human-agent teams, especially for systems such as MUM-T. Automation should ideally free operators from tedious, mundane, and time-consuming tasks; enabling them to focus on more critical responsibilities (Hart & Sheridan, 1984; National Research Council, 1982). However, automation does not completely remove all operational burdens from the human as it transitions the operator from a worker to a monitor, typically leaving the human responsible for the successful operation of the system. For instance, pilots controlling UAVs will usually be commanding and overseeing the actions performed by the UAVs and are likely to be held responsible incidents involving these aircraft as well as their own. The technology could become a distraction due to poor interface design, lag time, software bugs, user error, added stress, or unbalanced workload (Adams & Pew, 1990; Billings, 1991; Endsley, 1996; Hart & Sheridan, 1984; Norman, 1989). Even in normal flight operations, a majority of civilian pilots felt that automation increased workload due to manipulation and reprogramming requirements (Parasuraman & Riley, 1997; Parasuraman et al., 2000; Wiener, 1985, 1989). To prevent any tendencies towards these undesirable issues, the aim of a designer should be to identify the state at which the human remains in the control loop enough to attain situation awareness, but is not overexerted to the extent that performance deteriorates (Rusnock & Geiger, 2014).

Therefore, achieving the proper level of automation design and control between the pilot in the cockpit and the UAVs is a function of identifying where the pilot needs help. For this reason, it is crucial that researchers quantitatively capture the pilot's workload changes when operating UAVs to determine what an acceptable level of workload is such that the pilot is engaged and involved with flight tasks, but not oversaturated with responsibilities.

Mental Workload

Central to this research is the study of workload. Workload is a conceptual way to express the perceived demand experienced by a user in response to a specific task load (Beevis, 1992; Keller, 2002). Although most tasks have both a physical and cognitive component, the current research is primarily concerned with cognitive or mental workload. Wickens (2002) defined mental workload as "the relation between the (quantitative) demand for resources imposed by a task and the ability to supply those resources by the operator." For the purpose of this thesis, mental workload is defined as the relationship between an operator's mental capacity and the required attentional resources needed to perform a task at a given moment in time (Hart & Staveland, 1988). A person's capacity is a function of the following factors: environment, experience, level of training, proficiency, fatigue, stress, individual traits, and general workload strategy (Childress, M., Hart, S., & Bortolussi, 1982; Curry, Jex, Levison, & Stassen, 1979; Hart & Sheridan, 1984). Each of these factors contribute to the user's perceived mental effort, which can vary based upon the operator's aptitude to perform the task at hand.

Mental Workload and Performance

In past research, mental workload and performance have been studied together in an effort to explain the correlation between the two entities (Clare, Maere, & Cummings, 2012; Donmez B., Nehme C., & Cummings M.L., 2010; Hart & Staveland, 1988; Hebb, 1955; Reid &

Colle, 1988; Teigen, 1994; Yerkes & Dodson, 1908). Studies have found that mental workload generally increases as the number or complexity of user tasks increases and the time available to perform these tasks decreases (Hart & Staveland, 1988; Reid & Colle, 1988). However, the effect of workload on performance is not a linear relationship. Instead, performance will peak at a certain amount of workload before it begins to level off or decline (Teigen, 1994)**.** This relationship is often described by the Hebb-Yerkes-Dodson Law (Teigen, 1994; Yerkes & Dodson, 1908). The law describes the relationship of psychological arousal and performance as curvilinear for simpler tasks and an inverted-U for more difficult tasks. Figure 3 is as an adaptation of the Hebb-Yerkes-Dodson law with a simple and difficult task.

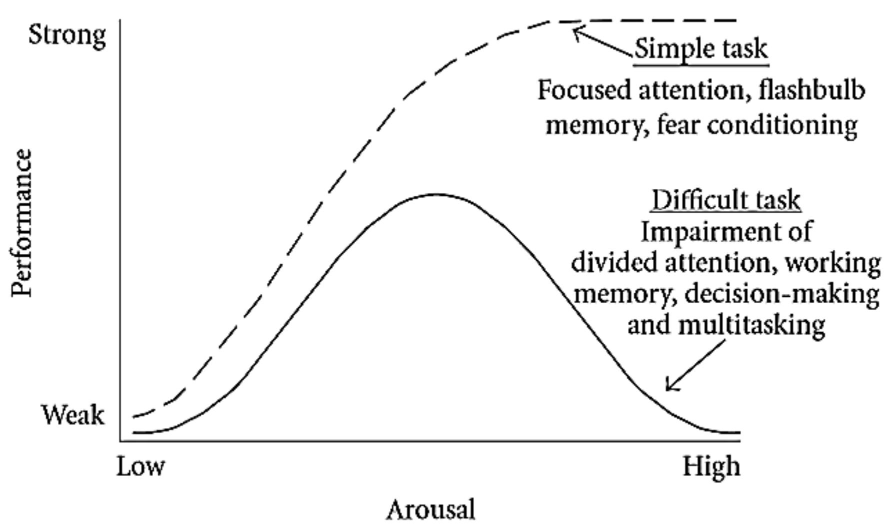


Figure 3. Depiction of the Hebb-Yerkes-Dodson Law– adapted from (Yerkes & Dodson, 1908)

The Hebb-Yerkes-Dodson law indicates that human performance increases with mental arousal, but only up to a point that is contingent upon the task complexity. Factors such as urgency, significance, and enjoyment all affect arousal level and can impact the person's attentiveness to a task. For both simple and difficult tasks, performance is poor when the human is unaroused (i.e. underloaded, unstimulated, or under-resourced) and generally increases as more operator resources (i.e. effort or focus) are invested in the task. For simple tasks, performance increases up to a certain level of arousal and then plateaus when the operator reaches his or her maximum level of cognitive capacity (Diamond, Campbell, Park, Halonen, & Zoladz, 2007). For more difficult tasks, performance increases with arousal, up to an optimal point after which the subject is over stimulated and performance is reduced as arousal increases (Hebb, 1955; Teigen, 1994; Yerkes & Dodson, 1908). Accordingly, maximum performance for complex tasks occurs at moderate levels of arousal because it permits the human to concentrate on relevant cues within the environment (Hebb, 1955; Teigen, 1994; Yerkes & Dodson, 1908). This relationship can also be extended to explain the impact of perceived mental workload on human performance. Mental workload has the same effect as psychological arousal, meaning that performance is degraded as the workload increases past the optimal point (De Waard, 1996; Wickens, 2008).

The correlation between perceived mental workload and performance can also be described by the Multiple Resource Theory (MRT) (Wickens 1984, 2002, 2008). According to Wickens (2002), the human operator has several different pools of mental resources that can be tapped simultaneously to process information. The multi-dimensional model is illustrated in Figure 4.

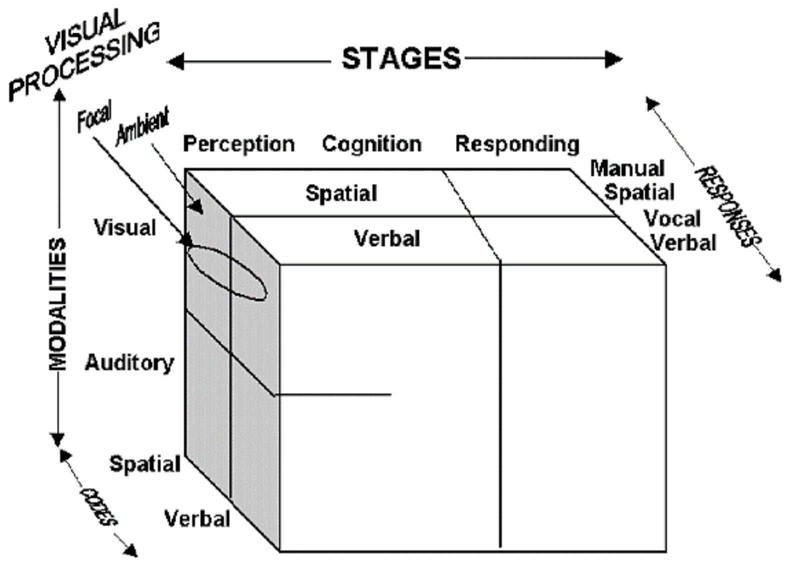


Figure 4. Multiple Resource Theory Model – adapted from (Wickens, 2002)

As Wickens explained, humans have a limited amount of cognitive resources, which restricts their ability to process information. The theory suggests that specific mental resources could be used in parallel, but the overuse of shared processing stages, perceptual modalities, visual channels, or processing codes could lead to resource interference and decreases in human performance (Wickens, 2002). For example, if a pair of tasks requires the same pool of cognitive resources (i.e. listening to two conversations at once), then the tasks must be handled sequentially because the auditory channel is overloaded with similar information. If a pair of tasks require different cognitive resources (i.e. scanning a crowd and listening to music), then the two tasks can be performed together because they do not stem from the same pool of resources within the brain. Furthermore, some tasks may require multiple resources, creating bottlenecks that limit parallel processing (Wickens, 2008). In either case, excess workload from a task demand can ultimately result in less efficient or less accurate performance from the operator.

In addition to the Hebb-Yerkes-Dodson law and MRT, other mental workload theories have been proposed to explain the relationship between workload and performance. Cassenti and Kelly (2006) proposed a workload curve, illustrated in Figure 5 with four regions: undertaxed, ceiling performance, steady decline in performance, and floor performance.

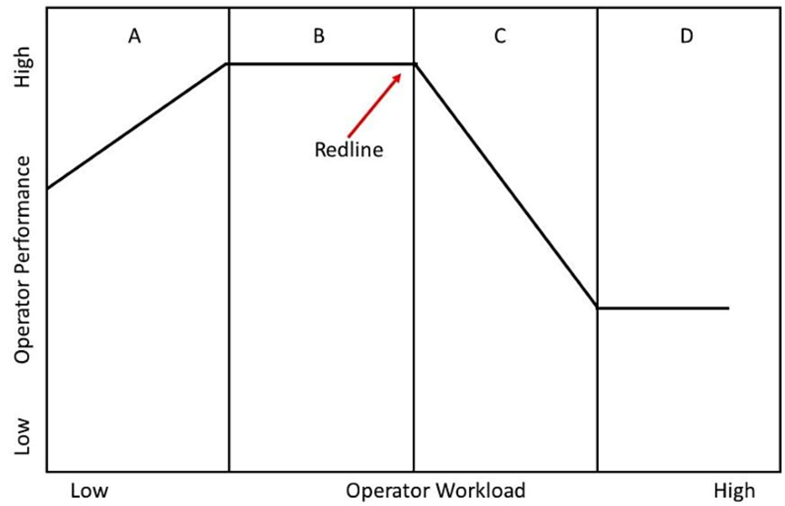


Figure 5. Operator Workload and Red-line – adapted from (Cassenti & Kelley, 2006)

Using this model, the level of workload resulting in maximum performance is described as an individual's red-line. The red-line occurs near the transition from region B to C as illustrated in Figure 5. Similar to the optimal point described in Figure 3, an operator's red-line is the maximum level of performance that an individual can sustain at the current task load before having to shed a task to continue functioning (Grier et al., 2008). If the workload exceeds the operator's red-line, then the individual will become overloaded and performance will deteriorate rapidly (Hart & Sheridan, 1984; Kahneman, 1973). On the other hand, operator workload that results in underload has also been shown to negatively impact task performance (Young & Stanton, 2002). Cognitive underload can occur when the operator is disengaged for an extended period of time, which can result in slower response speed and worsened precision (Hancock &

Chignell, 1988). In both circumstances, productivity or accuracy may diminish due to the overload or underload in workload

Based on the aforementioned studies, it is clear that the increase in workload can degrade performance as the pilot reaches cognitive saturation. Therefore, one would expect operators experiencing moderate levels of workload to perform better than those experiencing extreme levels of workload. By understanding the relationship between workload and performance, system designers can identify where the red-line of workload occurs and proactively decide the level of task load which is most acceptable for future improvements in human-agent teams. The ultimate objective is to reduce system complexity and enhance operator performance by leveraging automation where it can be most beneficial and appropriate.

Mental Workload and Expertise

Mental workload is also influenced by the level of information processing required by a specific operator. According to Neerincx (2003), there are three levels of cognitive information processing: automatic processes or skills, routine problem solving or rules, and more complex analysis of information. Experts or highly experienced operators may perform a task using automatic processing because they are more familiar with the system or task at hand. Conversely, novices or less experienced operators may need to spend more time, attention, or energy to perform a complex analysis of information so as to complete the same task (Hart & Sheridan, 1984; Secarea, 1990). Thus, the mental workload imposed by a given task load can vary significantly between individuals.

Mental Workload and Environment

Furthermore, the task load and the resulting perceived workload is not always constant during system operations. The task load and workload can vary due to changes in the environment, or some other external demand, which can influence the number and complexity of cues that an operator must process to correctly perceive the environment. Hollnagel and Woods' (2005) Extended Control Model can be used to understand how the cognitive demands of a task might change by revealing the dependencies among the layers of activities and simultaneous function of control loops. To start with, the model describes how the performance of a Joint Cognitive System takes place on several layers of control: tracking, regulating, monitoring, and targeting (Hollnagel & Woods, 2005). The demand of these simultaneous processes vary depending if the person is familiar with the task (i.e. novice vs expert), performs a task that requires several layers of control (i.e. flying and navigating), or encounters external issues (i.e. environmental disturbances). These variable factors can change the time constants and cognitive demands of a task, depending on their relative importance to the user's primary goals, consequently effecting the user's perceived level of workload and making workload difficult to model.

Measuring Mental Workload

Often varying significantly throughout a work period, workload can also be difficult to model perfectly because it cannot be directly observed. It must be inferred from the observation of overt behavior or the measurement of psychological and physiological processes (Cain, 2004). As a result, measuring human workload requires subjective testing based on the opinion of a participant or an expert or objective testing through computational approaches. In this thesis, the NASA Task Load Index (NASA-TLX), a subjective workload measure, and Visual, Auditory, Cognitive, Psychomotor (VACP) method, an objective workload measure, are covered (Bierbaum, Szabo, & Aldrich, 1989; Hart & Staveland, 1988).

Workload can be measured through both subjective and objective means. Subjective workload assessments are used to ask the participant to estimate the perceived mental workload they experience in response to a specific task load. They are frequently performed after an experiment is completed, typically using a survey or questionnaire such as the NASA-TLX (Hart & Staveland, 1988). These self-assessment evaluations (see Appendix 1) can capture the perception of mental workload, especially when the effects of many different contributing factors may be difficult to comprehend (Hart & Staveland, 1988). While subjective measures may easily lend themselves to both researchers and subjects, this type of measure can be influenced by an individual's personal judgment, heuristics, or biases. In many cases, subjective measures use a scaling system to record an individual's workload judgment about a task after the experiment is completed. However, information fidelity erodes as time elapses. If a task was performed early in the experiment or a questionnaire was conducted well after the task occurred, then the subject may only be able to accurately recall the most challenging or latest iteration of that task (Hart & Staveland, 1988).

The NASA-TLX is an example of a subjective workload evaluation technique that is pertinent to the research performed in this thesis. It was developed by the Human Performance Group at NASA's Ames Research Center over several years of laboratory studies involving simple manual control tasks, complex supervisory control tasks, and aircraft simulations (Hart & Staveland, 1988). The subjective assessment tool uses a multi-dimensional rating scale that measures the operator's perceived workload level by requiring subjects to rate task demands on six independent subscales: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration level (NASA, 1986). Each subscale is scored in five point increments on a 100 point scale and then prioritized from least to most important by the rater.

Descriptions of the six subscales are typically given in the form of questions and are shown in Table 6.

Table 6. NASA-TLX Subjective Measures – adapted from (Hart & Staveland, 1988)

Category	**Questions**
Mental Demand	How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
Physical Demand	How much physical activity was required? Was the task easy or demanding, slack or strenuous?
Temporal Demand	Temporal Demand: How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
Perceived Performance	How successful were you in performing the task? How satisfied were you with your performance?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

The overall workload score is then computed based on the weighted averages of all the subscales. Researchers can gain insight into how difficult a task is perceived to be and which resources are most important to the rater based on the task demand ratings selected for each subscale, the prioritization of subscales, and overall workload score. This method enables researchers to gain insight into the mental state of a human operator and the influence of task load on perceived workload with low intrusiveness and implementation requirements (Hart & Staveland, 1988). However, measuring mental workload through NASA-TLX scores has its disadvantages. First of all, a user may not recall their workload accurately because workload scores are reported after the task has been completed, rather than in the moment. Secondly, human-in-the-loop studies are time intensive and expensive. This makes it difficult for researchers to collect a large amount of data points in a short period of time. Finally, this method does not provide a way of measuring the second-by-second changes in workload that a user may experience in the course of a task.

On the other hand, objective measures have also been used to estimate an operator's mental workload. Objective workload measures are predictive in nature and can calculate the cumulative workload imposed by a series of tasks through computer simulations or direct performance measures. Given there is an established benchmark of tasks to be performed in a controlled environment, objective workload models can help researchers predict when an operator is near their red-line, identify which tasks are causing the red-line, and narrow in on which resource channel(s) are being overloaded (Bierbaum et al., 1989). For this reason, objective workload models can offer a better evaluation of workload throughout each stage of the system (D.K. Mitchell, 2000). This insight enables designers to pinpoint periods of high workload and modify the system design to mitigate burdening workload conditions for the operator.

One of the most reliable methods for modeling human workload is the Visual, Auditory, Cognitive, Psychomotor (VACP) method (Bierbaum et al., 1989) (see Appendix 2). Built upon Wicken's MRT (1984), the VACP model objectively assesses workload demands across the following seven resource channels: auditory perception, cognitive, fine psychomotor, gross psychomotor, speech, tactile, and visual perception. The VACP scale uses task ratings developed by Subject Matter Experts (SME) to explain the degree to which each resource component is used by a particular task over time (McCracken & Aldrich, 1984). Using this technique, VACP considers any excess demands placed on a specific resource by calculating the overall workload score for a particular instance in time for each VACP channel (Wickens, 2002). The fundamental idea is that tasks that utilize multiple resources will impose a higher workload on the operator because each VACP channel can only service one task demand at a time. In this manner the simulation model can offer predictive data on the VACP demands placed on an individual in a given scenario (Hugo & Gertman, 2012). Furthermore, the simulation model can also create a workload profile to display the changing workload demands over time in a given scenario.

In summary, findings from prior research reinforce the need for reliable measures of operator's mental workload when employing new systems such as MUM-T. It is important to not only utilize subjective workload measures, but also objective workload measures to quantitatively capture the operator's workload levels when performing a task or multiple tasks. Tools such as the VACP method offer greater insight into the pilot's workload changes when operating UAVs and can help system designers determine when the pilot is likely overloaded or underloaded and determine the responsibilities leading to the condition of concern.

Human Performance Modeling and IMPRINT

To integrate pilots and UAVs into a cohesive system, designers must consider the effect that Human-Agent Interactions (HAI) have on the pilot's cognitive workload. A useful tool for modeling cognitive workload and testing design options is through the Improved Performance Research Integration Tool (IMPRINT) (Alion Science and Technology Corporation, 2009). This section investigates the background and application of IMPRINT to explain how utilizing this tool is appropriate and useful for studying the effect of HAI on the pilot's cognitive workload when commanding UAVs.

Human performance modeling and simulation of operator workload are useful when trying to discover the innovative capabilities of new system designs and HAI with a system. In order to evaluate the workload that is imposed upon a pilot during air operations, engineers need a method to objectively measure the amount of workload produced within a given human-agent system. One approach to doing this is by performing a Human-In-The-Loop (HITL) study by building and testing multiple system designs and subjectively measuring the amount of workload experienced by each test subject. However, this process is inefficient and ineffective as it requires a simulation of each simulation condition to be constructed, recruiting and running the HITL simulations with multiple test subjects, and then analyzing and understanding the resulting data. Thus this approach can be time consuming and expensive. Furthermore, the workload values are specific to the test subjects and the simulated scenario.

An alternative workload measure would be to study the effect of HAI on the pilot's cognitive workload by using analytical modeling software. A modeling tool that could facilitate a method to estimate pilot workload is IMPRINT. IMPRINT can be utilized to create a Discrete Event Simulation (DES) that simulates the predicted workload of the pilot when interacting with both the cockpit and the UAVs. This alternative method shows promise in evaluating human workload during manned-unmanned flight operations because it is low cost and low risk.

Introduction of IMPRINT

IMPRINT is a dynamic, stochastic, discrete event modeling tool specifically designed to evaluate the interactions of human users and system technologies (Rusnock & Geiger, 2013). It was developed by ALION and funded by the U.S. Army Research Laboratory, Human Research & Engineering Directorate, to support manpower and personnel integration as well as human systems integration (Alion Science and Technology Corporation, 2009). Originally released in 1995, IMPRINT has been used with several human trial and theoretical experiments from human performance evaluations to the optimization of manning levels (Allender, 2000; Cassenti, Kelley, Colle, & McGregor, 2011; Harriott, Zhang, & Adams, 2013; Mitchell, D. K., Samms, C., & Wojcik, 2006; Rusnock & Geiger, 2014). It can test multiple alternative scenarios or system designs in a short period of time as well as quantify the mental workload for the human operator. In 2005, a more robust version of the program called IMPRINT Pro was released, which included tool upgrades as well as the ability to integrate the programming language C# for greater flexibility (Alion Science and Technology Corporation, 2009). However, for the purposes of the thesis, IMPRINT Pro will be referred to as IMPRINT.

Given this capability, modelers can use IMPRINT to analyze the cognitive demands experienced by the operators during specific tasks or at discrete time intervals throughout a scenario. The tool empowers modelers to discover emergent results in the data, test hypothetical adjustments of an interface, and determine the general efficiency of a system. Furthermore, this technique aids researchers in determining which tasks can be performed concurrently and which ones are likely to interfere with each other.

Fundamentals of IMPRINT

IMPRINT is a human performance modelling software that can be used to analytically study the effects of cognitive workload on operators during sample mission profiles. In this context, workload is defined as a measure of the task load, mental effort, or strain perceived by the human, with more tasks or more difficult tasks generally inflicting higher perceived workload. The theoretical basis for the mental workload option of IMPRINT is MRT where workload demands are assessed across several different resource pools to develop an objective measure of workload (Wickens, 2002). This enables researchers to account for demands placed on specific channels and identify any potential conflicts between them.

Using MicroSaint Sharp, an embedded discrete event task network modeling language, IMPRINT implements MRT by providing system designers with the ability to model human workload and performance as a function of time through tracked activities performed by the human or computer (Powers & Gacy, 2018). IMPRINT enables the system designer to use discrete task-level information to construct and parametrize task networks that represent the flow, performance time, and accuracy of operational missions. These task networks can be built by the system designer using either a VACP or advanced workload analysis. For the purpose of this research, only the VACP method will be discussed in further detail as it is the most appropriate method for analyzing the MUM-T system.

IMPRINT also consists of four autonomous modules: the Equipment, Warfighter, Forces, and Mission modules. Each module is purposely designed to offer specific data outputs to inform different decisions (Alion Science and Technology Corporation, 2009). For the purpose of this research, specific focus will be placed on the Mission module. The Mission module, using the VACP analysis method, simulates the effects of task times and workload ratings for each resource on the overall system performance. The task networks in this type of module are developed using direct observations and data collection to estimate task time probability distributions for each action and mental workload values for each human operator action. Furthermore, the various system allocations can be modeled and manipulated to incorporate automation by assigning specific tasks to be performed by the human or machine.

During the mission module simulation, IMPRINT predicts task performance and calculates how much workload each operator is experiencing throughout the mission (Alion Science and Technology Corporation, 2009). When using the VACP workload methodology in the Mission module, system designers must identify

1) The tasks necessary to operate a proposed system,
2) The order or logical conditions in which they must be performed,
3) The distribution of time duration for performing the tasks,

4) The operators who perform them, as well as
5) The workload and the mental resources expected to be used for each task (Hamilton & Bierbaum, 1992)

This is accomplished by first completing a task analysis. A task analysis outlines the sequence of tasks performed, timing of the tasks, workload associated with each task, and allocates these activities to the human or computer. This information is used to develop a task network in IMPRINT. Each task in the network is assigned a workload rating from one to seven for each of the following VACP channels: visual, auditory, cognitive, fine motor, gross motor, speech, tactile, or a combination of any of these seven resources. See Appendix 2 for the standardized VACP values used in IMPRINT (Alion Science and Technology Corporation, 2009). The workload ratings are combined in the IMPRINT simulation to create a workload profile that provides an objective measure of workload at each instant throughout the trial. This data also makes it possible for researchers to calculate a time-weighted average workload across the entire trial. Using this information, it is possible to show the relationship between workload and performance.

Once the baseline task network has been developed, small changes can be made to the task flow to test several design concepts. For instance, the task flow can be executed several times using variations in the task times or frequency of occurrence to assess different goals and operator workload levels. In addition, IMPRINT can be customized by the system designer who can write C# code to perform specific actions at certain times, such as the beginning or ending of specific tasks. By providing a blend of pre-structured tools and programming flexibility to the modeler, the resulting data from these simulations can be further analyzed to determine the effects of activities on human workload.

At the completion of the simulation, IMPRINT can compare the minimum acceptable mission performance time and accuracy to the predicted performance. Using these results, the modeler can study the range of outcomes that occur in the mission. This is a valuable capability for analyzing workload data because it can help analysts determine whether an operator such as a pilot is task saturated when performing specific activities such, as commanding one or multiple UAVs.

Research Gap

As a whole, IMPRINT is a valuable tool for defining the operators and the workload of tasks, providing an automated means of task switching, and generating reports that highlight the results of both system and human performance variation. It can test multiple alternate scenarios in short amounts of time, consuming fewer resources than a HITL experiment. These capabilities are what make IMPRINT a powerful and effective method for modeling the effect of HAI on an operator's cognitive workload when commanding a machine or computer agent.

Historically, researchers have predicted mental workload using IMPRINT to address complex models concerning system design and human behavior interactions. IMPRINT has been previously used to perform human workload modeling for multiple human-agent technologies such as Shadow UAVs (Hunn & Heuckeroth, 2006), Micro Air Vehicles (Pomranky & Wojciechowski, 2007), U.S. Army Tanks (D. K. Mitchell, 2009), and autonomous ground vehicles (Pop, Michelson, & Engineering, 2018). It has also been used to evaluate mental workload differences between human-human teams versus human-robot teams (Harriott, Zhuang, Adams, & Deloach, 2012), and determine manpower requirements for military applications and research (Rusnock & Geiger, 2013). Each of these examples illuminate the wide range use of IMPRINT technology for modeling human-machine interactions.

In a 2011 study conducted by Schneider and McGrogan, used IMPRINT to model the potential effects of Multi-Aircraft Control (MAC) on pilot workload when implementing MAC with the MQ-1B Predator system architecture. This research concluded that pilots experienced low workload when operating one or two Unmanned Aircraft Systems (UAS) during benign operations. However workload quickly built up when pilots operated three or more UASs and became unmanageable for a single pilot to handle during dynamic operations. This study highlighted the need for techniques and technology to reduce task and communications demands on UAS pilots to effectively implement MAC.

While MAC for UAS have been studied, there is limited research investigating workload impacts of MUM-T in military flight operations. This is likely due to the novelty of this type of technology and human-agent system integration. In view of that, this thesis used IMPRINT to gauge the effect of HAI on the pilot's cognitive workload when commanding UAVs. Similar to the research conducted by Schnieder and McGrogan (2011), it is possible to model existing operation procedures and inputs from external stimuli (i.e. UAVs flying around a fighter aircraft) to observe and predict workload levels through IMPRINT. The MUM-T system has been broken down into human and autonomous components where the operator (i.e. pilot) and the agents (i.e. UAVs) can accomplish a measurable set of finite tasks that are assigned corresponding workload values. Using this task network, IMPRINT can help system designers explore the effects of MUM-T on the pilot's cognitive workload and the human-agent team's mission performance.

Summary

The literature presented in this chapter builds the necessary background knowledge to understand the research performed in the study and the overall significance of this work. The chapter focused on the development of automation, the concept of workload, and the relationship between the two with regard to human performance modelling. This research aims to gather each of these research concepts together to develop a cohesive study that investigates how IMPRINT can be applied in a novel way to quantitatively model the mental workload of pilots when they are operating their aircraft and commanding UAVs simultaneously.

Understanding workload theory and the application of IMPRINT will enable the reader to answer the first two investigative questions regarding what effect(s) HAI have on the pilot's cognitive workload and overall mission performance when commanding UAVs. The final research question focuses on different amounts of autonomous control abstraction. It investigates how much of the operator's cognitive tasks should be relinquished by the command pilot and reassigned to the UAVs to reduce the amount of workload experienced by the pilot to reduce operator workload and increase mission performance in the flight operation task. Finally, the research around this topic was explained, demonstrating a gap that needed to be filled and how this study aims to contribute to the body of knowledge by using IMPRINT in an innovative way.

III. A New Model of Airpower: Development of an IMPRINT Model to Analyze the Effects of Manned-Unmanned Teaming on Mental Workload

Abstract

Due to the advent of autonomous technology coupled with the extreme expense of manned aircraft, The Department of Defense (DoD) has increased interest in developing affordable, expendable Unmanned Aerial Vehicles (UAVs) in Manned-Unmanned Teaming (MUM-T). This concept employs UAVs to become autonomous wingmen for jet fighters in mosaic warfare. With a single pilot commanding the UAVs while piloting their aircraft, they may find it challenging to manage all systems should the system design not be conducive to a steady state level of workload.

To understand the potential effects of MUM-T on the pilot's cognitive workload, an Improved Performance Research Integration Tool (IMPRINT) Pro, pilot workload model was developed. The model predicts pilot workload in a simulated environment when interacting with the cockpit and multiple UAVs to provide insight into the effect of Human-Agent Interactions (HAI) on the pilot's cognitive workload and mission performance. This research concluded that peaks in workload occur for the pilot during periods of high communications load and this communication may be degraded or delayed during air-to-air engagements.

Key Words

Human-Agent Interactions, Unmanned Aerial Vehicles, Manned-Unmanned Teaming, Mental Workload, Improved Performance Research Integration Tool, Human Performance Modeling

Introduction

Due to the advent of autonomous technology coupled with the extreme expense of manned aircraft, the DoD has developed an interest in constructing affordable Unmanned Aerial Vehicles (UAVs) to become autonomous wingmen for jet fighters in mosaic warfare (Drew, 2016). Like a mosaic that forms a whole picture out of smaller pieces, battlefield commanders can utilize disaggregated capabilities, such as low-cost UAVs, to operate in contested environments (Magnuson, 2018). Utilizing UAVs to complement manned aircraft may offer advantages such as increased pilot survivability as well as amplified firing power to fill capability and capacity gaps. However, there are complications with this new strategy. For example, in an envisioned architecture, commonly referred to as Manned-Unmanned Teaming (MUM-T), command pilots will need to deploy capabilities from the UAVs in addition to controlling their own aircraft. The need to devote attention and mental resources to both controlling their own aircraft and the UAVs could be challenging for pilots should the system interface design not be conducive to maintaining a manageable level of workload.

To integrate pilots and UAVs into a cohesive system, designers must consider the effect that Human-Agent Interactions (HAI) have on the pilot's cognitive workload. In this context, workload is defined as a measure of the task load, mental effort, or strain perceived by the human, with more tasks or more difficult tasks generally inflicting higher perceived workload. To evaluate the workload that is imposed upon a pilot during air operations, engineers need a method to objectively determine the amount of workload produced within a given human-agent system. One approach is to perform Human-In-The-Loop (HITL) experimentation by prototyping and testing multiple system designs, including subjectively measuring the workload experienced by test pilots who fly simulated missions within the prototype system. While human research and prototyping of automation produces valuable information, it is inefficient and ineffective as the process is tedious, lengthy, and costly to complete. There can also be safety issues involved when performing risky HITL experiments. As such, to design a system using this approach as the only feedback mechanism constrains the number and variety of alternative system designs which can reasonably be considered within a design effort.

An alternative to HITL evaluations is to assess cognitive workload through analytical modeling. A modeling tool that could be employed to estimate pilot workload is the Improved Performance Research Integration Tool (IMPRINT). IMPRINT quantitatively models operator workload across several different resource channels through the incorporation of the Visual, Auditory, Cognitive, and Psychomotor (VACP) scale (see Appendix 2) (Bierbaum et al., 1989). The tool can be used to simulate various system configurations and their effects on pilot workload within a Discrete Event Simulation (DES). This method can provide a lower cost method than HITL evaluations and permit a greater number of alternative design options to be explored. This tool can be particularly effective when coupled with HITL evaluations to provide validation and to ground assumptions about human behavior in novel circumstances, where human behavior is often unpredictable (Goodman, Miller, Rusnock, & Bindewald, 2017; Rosenberg, 1982).

In our current research, IMPRINT was used to construct a DES to assess the effects of MUM-T on operator cognitive workload and system performance. The baseline DES represented tasks performed by human subjects enrolled in a previously conducted HITL evaluation (Schumacher et al., 2017). The study replicated a dynamic, military, offensive counter-air scenario in which individual performance and mental workload could vary in real-time based on the operators' capabilities.

An original baseline DES was developed to quantitatively capture the pilot's cognitive workload levels when controlling both UAVs and manned aircraft. An alternative system configuration was then created to compare the baseline model to traditional aviation techniques. The findings presented in this research provided a significant step towards simulating the complexities of real-world activities by mirroring the highly dynamic nature of realistic military operations in a simulated environment.

Method

Design of the ATACM Study

In order to understand this data set, the participants, mission scenario, and task environment from this study is reviewed in this section. Nine experienced former military pilots participated in the Autonomy for Air Combat Missions (ATACM) study. The ATACM study was a HITL experiment that developed and tested critical autonomous decision and machine learning technologies in a virtual simulation cockpit with the aim of enabling a single pilot to command multiple UAVs in flight while controlling his or her own aircraft in highly contested environments (Schumacher et al., 2017). After initial training and practice, each pilot flew four air-to-air trial engagements in which the pilot commanded three UAVs against four adversaries. For each trial, participants were given ten minutes to employ their own aircraft and those of the UAVs to destroy the four adversaries before the push point was reached. The scenario ended when any of the following occurred: 1) all four adversary aircraft were killed, 2) all three UAVs or "wingmen" were killed, 3) the pilot was killed, or 4) the push point was reached at ten minutes. The general mission scenarios are illustrated in Figure 6.

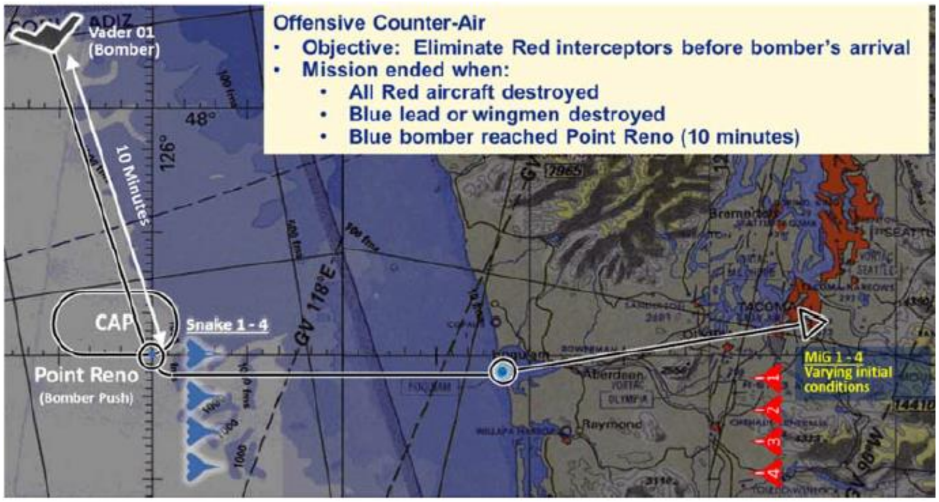


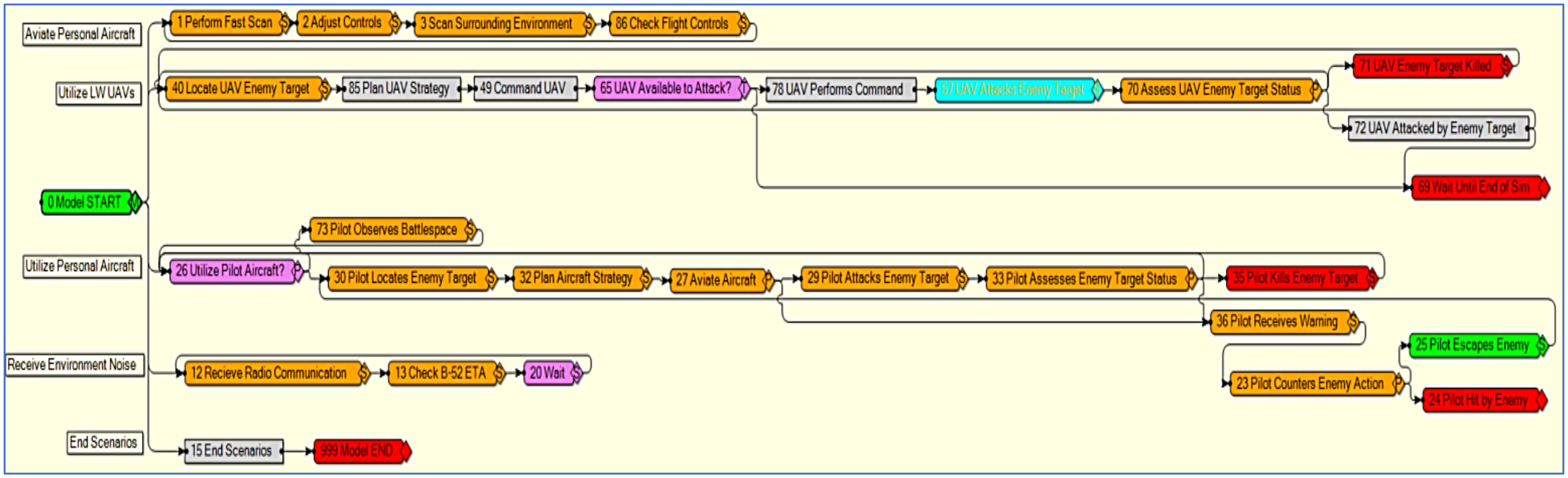
Figure 6. ATACM Mission Scenarios (Schumacher et al., 2017)

The virtual simulation cockpit utilized in the ATACM experiment was composed of four major elements: 1) a pilot-vehicle interface, 2) a multi-UAV artificial-intelligence-based multi-agent controller, 3) automated (scripted) low-level responses to commands, and 4) a virtual piloted mission simulation. Using these four resources, the test subjects were required to locate and target adversary aircraft by commanding three UAVs and utilizing their own aircraft to fire at targets. Video footage from the experiments was captured and used for analysis in this research.

IMPRINT Baseline Model Development

The information provided from the HITL was used to create the baseline DES model for a single human pilot commanded three UAVs against four enemy targets (see Appendix 3). As shown in Figure 7, the baseline task network model was composed of four task loops and one logic loop: 1) Aviate Personal Aircraft, 2) Utilize UAVs, 3) Utilize Personal Aircraft, 4) Receive Environment Noise, and 5) End Scenarios.

1) **Aviate Personal Aircraft**: the first task loop included basic tasks such as adjusting the flight controls or scanning the surrounding environment that the pilot performed when operating his or her own aircraft.
2) **Utilize UAVs**: the second task loop included tasks such as commanding the UAV or supervising UAV attacks, which the pilot executed to deploy the UAVs. The pilot commanded the UAVs using a continuum of autonomous control abstraction that ranged from simple commands such as “turn left” or “fly at an altitude” to more complex commands such as “fly formation” or “attack target.”
3) **Utilize Personal Aircraft**: the third task loop included tasks such as aviating the manned aircraft or attacking the adversary target, which the pilot performed in order to utilize his or her own aircraft to attack the enemy.
4) **Receive Environment Noise**: the fourth task loop included the workload associated with receiving audio notifications over the radio.
5) **End Scenarios**: the final logic loop included tasks that would trigger the DES to end if any of the stopping scenarios were fulfilled



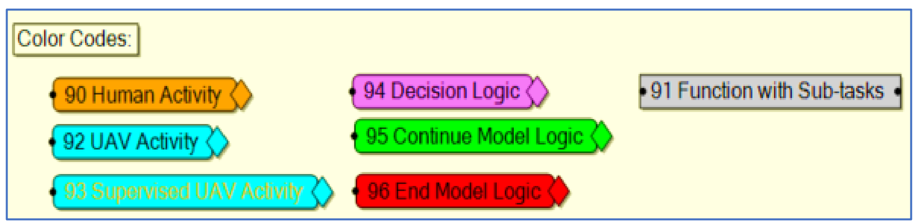


Figure 7. Baseline IMPRINT Task Network

Each of these task loops ran in parallel with one another as it was assumed that the pilot performed these activities concurrently. The final logic loop also ran concurrently with the other task loops in order for the software to evaluate whether or not the simulation satisfied one of the ending conditions. Once the task network was developed, each task was assigned a VACP workload value, task time, and decision probability. The finalized model was then validated in comparison to results obtained from the ATACM study (see Appendix 3).

Within the DES, the independent variable was the use of UAVs in the DES. The dependent variables were the mission performance and mental workload of the pilot during a simulation run. In the first model set up, both the manned aircraft and UAVs were employed to attack the adversaries. In the second model set up, only the manned aircraft was employed to attack the adversaries. The mission performance was measured by calculating the number of enemy targets that survived. The workload of the pilot was determined using the VACP scores gathered from each model for a subset of thirty trials, producing a time- average for the baseline model.

Analysis and Results

After the creation of the baseline model, one thousand DES trials were run to study the effect of HAI on the pilot’s cognitive workload when commanding three UAVs against four enemy targets. In the first “baseline” model setup, both the manned aircraft and UAVs were employed to attack the adversaries. In the second “manned-only” model set up, only the manned aircraft was employed to attack the adversaries. For each condition, the mission performance and mental workload of the pilot were calculated and then analyzed to compare how the system was effected by the incorporation of MUM-T.

Mission Performance Analysis

Figure 8 shows the percent of trials as a function of the number of enemy targets remaining at the end of each trial.

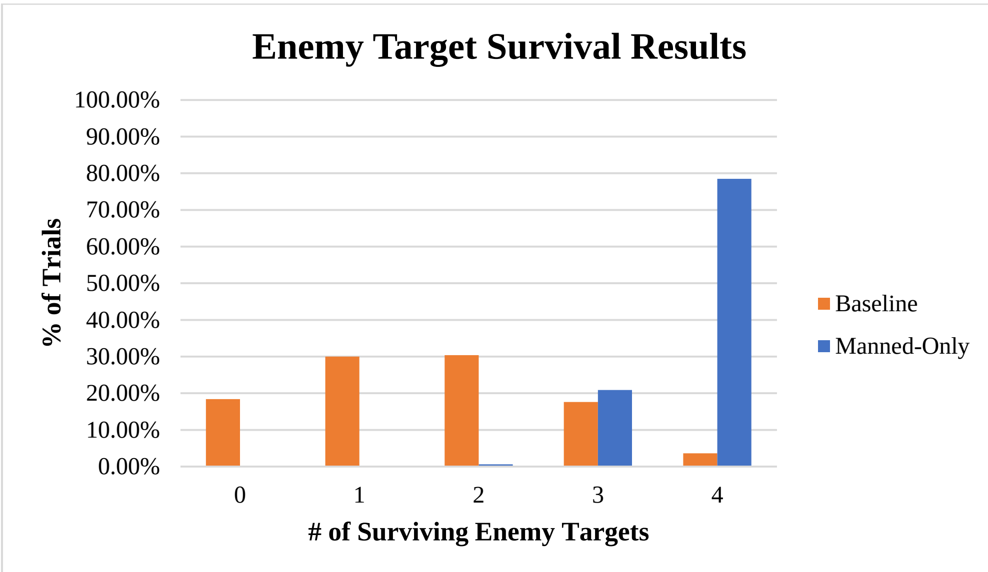


Figure 8. Graph of Enemy Target Survival Results

According to the data, the number of surviving enemy targets was reduced when the UAVs were incorporated into the model. The manned-only condition had 4 enemy targets survive per trial on average, while the baseline condition only had 2 enemy targets survive per trial on average. Furthermore, the incorporation of the UAVs resulted in all of the enemy targets being killed in 18.40% of the simulation trials. Conversely, 0% of the simulation trials resulted in all of the enemy targets being killed in the manned-only condition. This significant difference was expected due to the added attack capability that the pilot had with the three UAVs attacking four enemy targets instead of a single pilot carrying the weight of the battle. For this reason, the incorporation of UAVs improved the human-agent team's mission performance. Despite this result, the workload levels must also be analyzed to determine whether or not the pilot would be oversaturated with tasks when utilizing this supplementary technology. It is important to look at this difference in workload against the increase in mission capability to determine whether changes in workload levels are worth the improvement in mission performance.

Workload Profile Analysis

In this section, the total objective workload experienced by the operator was compared between the UAVs and manned-only DES models. IMPRINT calculated a workload summary based on the length of time the pilot spent performing a specific activity in relation to the combined VACP value(s) assigned for the interfaces of each task node. Events that were above a workload level of 60 were considered to be near or above the saturation threshold where the system imposed more work than the pilot could effectively perform (Mitchell, 2003; Schneider & McGrogan, 2011). In an ideal mission scenario, all workload levels would be below 60.

It should be noted the NASA-TLX (Hart & Staveland, 1988), a self-assessment workload survey (see Appendix 1), was utilized in the ATACM study to record the test subjects' individual workload judgments about a task after the experiment was completed. It was not possible to perform an Analysis of Variance (ANOVA) test to compare the objective workload values obtained from IMPRINT and subjective workload values obtained from the NASA-TLX surveys because only one condition was used from the ATACM study. Thus there is no variation expected in the NASA-TLX values. For this reason, only an analysis of the IMPRINT workload profile could be performed.

The workload graph shown in Figure 9 provided insight into some of the interactions and implications from incorporating MUM-T into flight operations. At the beginning of the simulation, the VACP value during the first part of the profile varied from 32 and 46 as the pilot planned the attack and deployed the UAVs in addition to his or her own aircraft to track the enemy targets. In the next phase, the workload consistently fluctuated between 40 and 42 when the pilot navigated the aircraft and supervised the UAV activity. This moderate level of workload was well below the saturation threshold, which suggested that these activities were manageable for the pilot as long as the aircraft did not experience any emergencies.

The attack began in the third phase, causing the workload to spike above the red-line to a maximum of 61 when the pilot needed to scan the surrounding environment, assess the enemy target’s status, navigate the aircraft, and receive radio communications. It slowly declined to a minimum workload level of 32 when the attack subsided. Then the workload resumed to a manageable and steady pattern when the pilot subsequently returned to navigating the aircraft and supervising the UAVs in the fourth phase. However, this manageable level of workload did not last long. The mean workload immediately increased above the saturation threshold in the fifth phase when the pilot received radio communications for the second time and then slowly declined once again. The sharp spikes in workload indicated that the incorporation of communications is a failure point. The workload level is generally manageable, but it will require the pilot to employ workload mitigation strategies when communicating with other aircraft beyond the UAVs.

In the sixth phase, the pilot returned to supervising the UAVs and navigating the manned aircraft. For an instant, the pilot experienced a sharp spike to 51 in workload due to the pilot receiving radio communication and supervising the UAVs to attack an enemy target at the same time. Despite this spike and slight workload fluctuations in phase seven, the workload levels indicate an ideal situation for human-agent teaming with all of the aircraft in a benign mission mode.

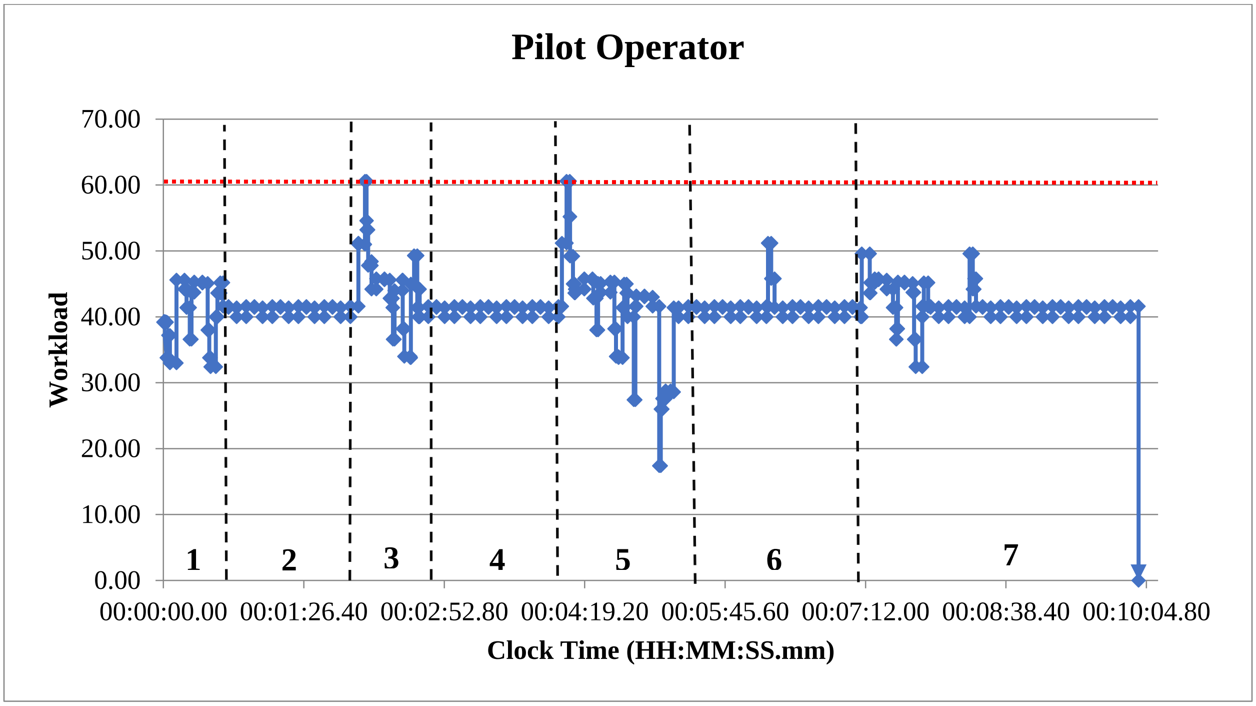


Figure 9. IMPRINT Workload Profile for Pilot in Baseline Model

The workload graph shown in Figure 10 provided insight into some of the interactions and implications when MUM-T is eliminated from flight operations. With the exception of commanding any UAVs, the pilot performed the same tasks as described in the analysis of the baseline model workload profile. At the beginning of the simulation, the VACP values over the first part of the profile generally varied from 32-34 as the pilot planned the attack and deployed his or her own aircraft to track the enemy targets. In the next phase, the workload momentarily spiked in two instances when radio communication was received. Despite these cases, the workload consistently fluctuated from 32-34 as the pilot performed aircraft navigation and control. Even with the slight uptick in workload, the level of workload experienced by the pilot was well below the saturation threshold. This reasonable level of workload suggested that basic aircraft control and navigation activities with no enemy engagement are manageable for the pilot.

The attack began in the third phase, causing the workload to spike to a maximum of 47 when the pilot needed to use the aircraft to attack the enemy target and receive radio communications. It steadily declined to a minimum of 18 when the attack subsided and pilot resumed normal aircraft navigation and control in the fourth phase. Despite the slight spike to 42 in workload due to the transmission of radio communication, the workload levels were generally stable for the remainder of the mission. Throughout the mission, the pilot's workload was manageable and much lower than the workload experienced in the DES including MUM-T. This was expected considering the pilot only needed to focus on his or her aircraft and did not need to command three other UAVs in addition to the manned plane.

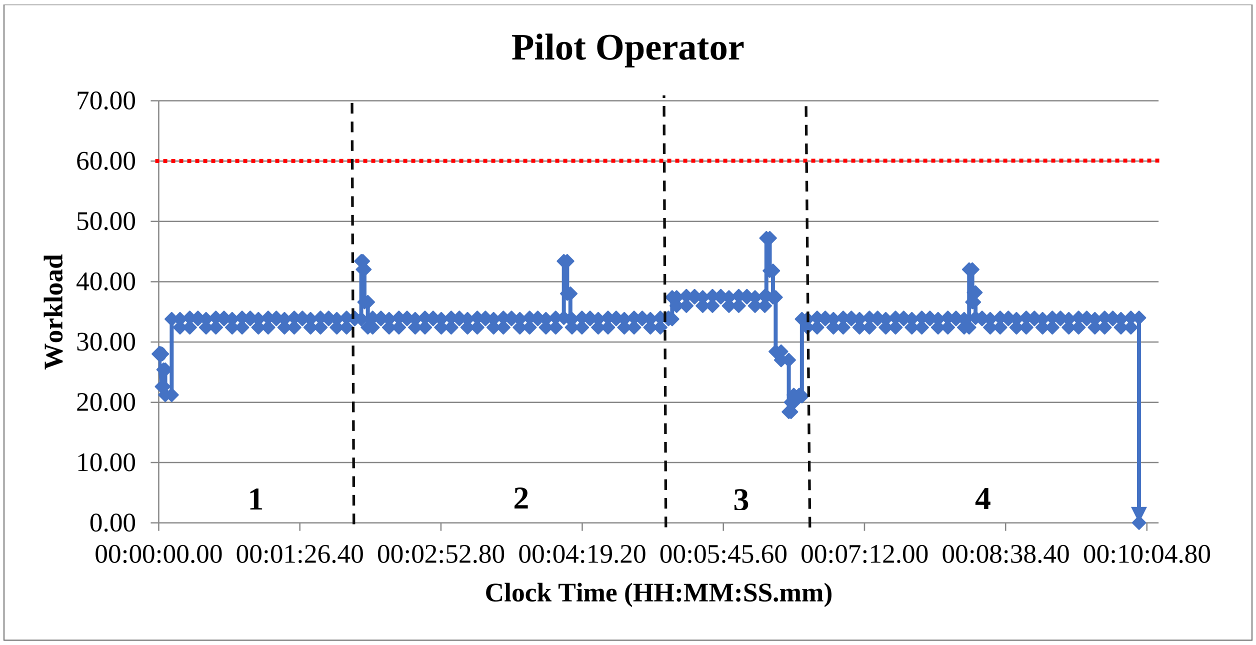


Figure 10. IMPRINT Workload Profile for Pilot in Manned-Only Model

Time-Persistent Average Workload Analysis

Using the VACP workload values from IMPRINT, a single representative workload value was also computed by taking the time-persistent average across 30 DES trials. The time-persistent average illustrated how hard the pilot worked as a whole to command the three UAVs by weighting the workload values by the duration the workload was experienced. According to the data, the pilot experienced a time-persistent average workload value of 42.34 for the baseline model. On the other hand, the pilot experienced a time-persistent average workload value of 33.83 for the model lacking MUM-T. The results indicated that the pilot's cognitive workload was mostly below the saturation level for both scenarios, but it varied significantly throughout the simulations.

Through an analysis of the mission performance, workload profiles, and time-persistent averages, it was determined that the increase in mission capability is worth the difference in the pilot's cognitive workload levels. The incorporation of MUM-T in flight operations improved the pilot's ability to successfully strike enemy targets and was manageable as long as the pilot did not require immediate attention for anything critical such as aircraft emergencies or prolonged external communication. In the simulation setup, both the manned and unmanned aircraft were utilized to attack four enemy targets. There were two moments in time when the threshold saturation of 60 was exceeded due to incoming radio transmissions. However, these spikes were infrequent and most of the workload was well below the saturation threshold. This suggested that the operator workload is manageable for the pilot with some communications offloading, when necessary. In the event of higher levels of radio communications, which are likely in operational air missions, workload mitigation strategies will be required to ensure that there is no mission degradation.

Conclusion

The research performed in this study sought to use DES to understand the effects of HAI on the pilot's cognitive workload when commanding UAVs. This was accomplished by examining the tasks performed by human subjects in the ATACM study, and then designing a simulated task environment modeled after these tasks. The model was built in IMPRINT to investigate how human cognitive workload and mission performance was impacted when a pilot commanded three UAVs in addition to his or her personal aircraft. The DES was validated by comparing the mission performance and timing results to that of the ATACM study. The results of the simulation indicated that mission performance was improved by the use of 3 UAVs against 4 enemy targets in an air-to-air operation. Furthermore, peaks in workload occurred for the pilot during periods of high communications load and this communication may be degraded or delayed during air-to-air engagements. Using this information, designers could predict potential workload issues when the pilots command the UAVs and communicate with other aircraft or ground stations in future MUM-T systems.

Future work in this area of research includes additional examination of alternative scenarios. In the next study, an alternate model will be created to simulate varying levels of autonomy to determine what would be the optimum level for operation of multiple UAVs. Furthermore, the current research is limited to data provided by the ATACM experiment. The next step would be to gather data that exists outside of a HITL experiment to develop a model that more realistically captures HAI between pilots and their UAVs in an operational environment. Once this type of data becomes available, an improved model could be used to investigate improvements in MUM-T without the problematic costs of time, money, and resources.

IV. Simulation-Based Evaluation of the Effects of Varying Degrees of Control Abstraction for Manned-Unmanned Teaming on Mental Workload of Pilots

Abstract

The future of air combat is expected to evolve significantly to include new technologies and novel concepts of operation. The Manned-Unmanned Teaming (MUM-T) concept involves low cost, attritable Unmanned Aerial Vehicles (UAVs) that could be deployed alongside a manned aircraft. The UAVs act as a complementary asset and bolster offensive air operations. Given the complexity of future operating environments, the degree of autonomous control required for pilots to concurrently operate multiple UAVs and their own aircraft is one area of concern. To determine the amount of autonomous control abstraction that has the largest impact in reducing operator workload and increasing system performance, a predictive workload model was developed using the Improved Performance Research Integration Tool (IMPRINT). This research concluded that the UAVs should be commanded through a combination of Vector Steering, Pilot Directed Engagements, and Tactical Battle Management to increase mission performance and maintain the pilot's cognitive workload at a manageable level.

Key Words

Human-Agent Interactions, Unmanned Aerial Vehicles, Manned-Unmanned Teaming (MUM-T), Mental Workload, Improved Performance Research Integration Tool, Human Performance Modeling, Level of Automation, Autonomous Command and Control

Introduction

The U.S. Air Force's 2016 Small Unmanned Aircraft Systems Flight Plan (Secretary of the Air Force Public Affairs, 2016) described the long-term vision for remotely piloted aircraft in the next 20 years. It was envisioned that a single operator would command multiple platforms of

Unmanned Aerial Vehicles (UAVs), such as in Manned-Unmanned Teaming (MUM-T) system, where one or a few full-sized remote aircraft would take on traditional manned wingman roles. To fulfill this vision, a surge of developments have been made in the development of MUM-T platforms. Several prototypes of MUM-T have taken flight, most notably the Air Force Research Laboratory's (AFRL) XQ-58A Valkyrie wingman. The XQ-58A is a long-range, high subsonic UAV which completed its first flight in March 2019 (88 Air Base Wing Public Affairs, 2019). The successful completion of this experimental flight test is a major step forward towards integrating small robotic fighter jets into air warfare. However, the United States is not the only country dabbling in such technology. Other countries such as Australia and China have already started to develop increasingly sophisticated UAVs to supplement their military's air operations (Joe, 2019; Stevenson, 2019).

With the Pentagon's increasing focus on competing with China and Russia for military dominance, the Department of Defense (DoD) must re-evaluate some of the basic conventions of flying to leverage the best of traditional aviation and emerging capabilities to maintain its dominance in the skies. It is expected that American UAVs, similar to the likes of the XQ-58A, will be paired with an F-22 Raptor or F-35A Joint Strike Fighter to give the United States Air Force's two stealth fighters the ability to fight in combat like never before (Tevithick, 2019). A single fighter aircraft could have several UAVs, each carrying additional weapons, radars and communication data links. The platform would increase pilot survivability by scouting ahead, absorbing enemy fire, and multiplying the enemy's targeting. Additionally, these resources could give the command aircraft amplified firing power.

However, there are complications with this new strategy should the DoD choose to adopt MUM-T for frontline use. While pilots have controlled UAVs from afar using Remotely Piloted

Aircraft (RPA), the idea of flying both manned and unmanned aircraft presents a bigger training challenge (Wassmuth & Blair, 2018). The command pilot bears the weight of the combat effort and will need to deploy capabilities from the UAV in addition to controlling the manned aircraft. The challenge of maintaining close control of UAVs requires a new approach to autonomous control and integration. This balancing act could be difficult for pilots to maintain should the system interface design not be conducive to maintaining a manageable level of workload.

Therefore, system designers must understand the potential effects of varying amounts of autonomous control when designing human-agent teams, especially for systems such as MUM-T. Automation should ideally free operators from tedious, mundane, and time-consuming tasks, enabling them to focus on more critical responsibilities (Hart & Sheridan, 1984; National Research Council, 1982). However, automation does not completely remove all operational burdens from the human as it transitions the operator from a worker to a monitor. For instance, pilots controlling UAVs will usually be commanding and overseeing the actions performed by the UAVs. The technology could distract the pilot from managing the battle or flying their own aircraft due to poor interface design, lag time, software bugs, user error, added stress, or an unbalanced workload (Adams & Pew, 1990; Billings, 1991; Endsley, 1996; Hart & Sheridan, 1984; Norman, 1989).

To prevent any tendencies towards these undesirable issues, it is crucial that researchers investigate how people and automation can effectively team to give the operators a level of workload which permits them to perform time critical control tasks. Previous research has provided insight into framing the amount of control abstraction between the human operator and the agent based on the level of control inputs given by the operator and interdependencies between the two (C. D. Johnson et al., 2017; M. Johnson, Vignati, et al., 2018). The five Levels of Human Control Abstraction (LHCA) (C. D. Johnson et al., 2017) describes the cognitive tasks relinquished by the human operator and reassigned to the automation. The reduction of operator control inputs can be modelled using the Interdependence Analysis Tool (IAT) (M. Johnson, Vignati, et al., 2018). This tool helps designers to visually see how human operators and automation support one another in a joint activity. Through an analysis of the effects of increasing autonomous control on the pilot's cognitive workload and mission performance, this research uses both of these frameworks to provide a recommendation for selecting potential system designs for the interactions between the pilot in the cockpit and the UAVs in the sky. This is a significant area to explore because the command structure of the overall platform affects the human's cognitive workload and, thus, the human-agent team's overall mission performance in combat.

Method

The main objective of this research was to evaluate how much of the operator's cognitive tasks should be relinquished by the command pilot and reassigned to the UAVs to reduce operator workload and increase mission performance in the flight operation task. The IMPRINT model illustrated in Figure 11 (Andrews, 2020) was modified to address this research question.

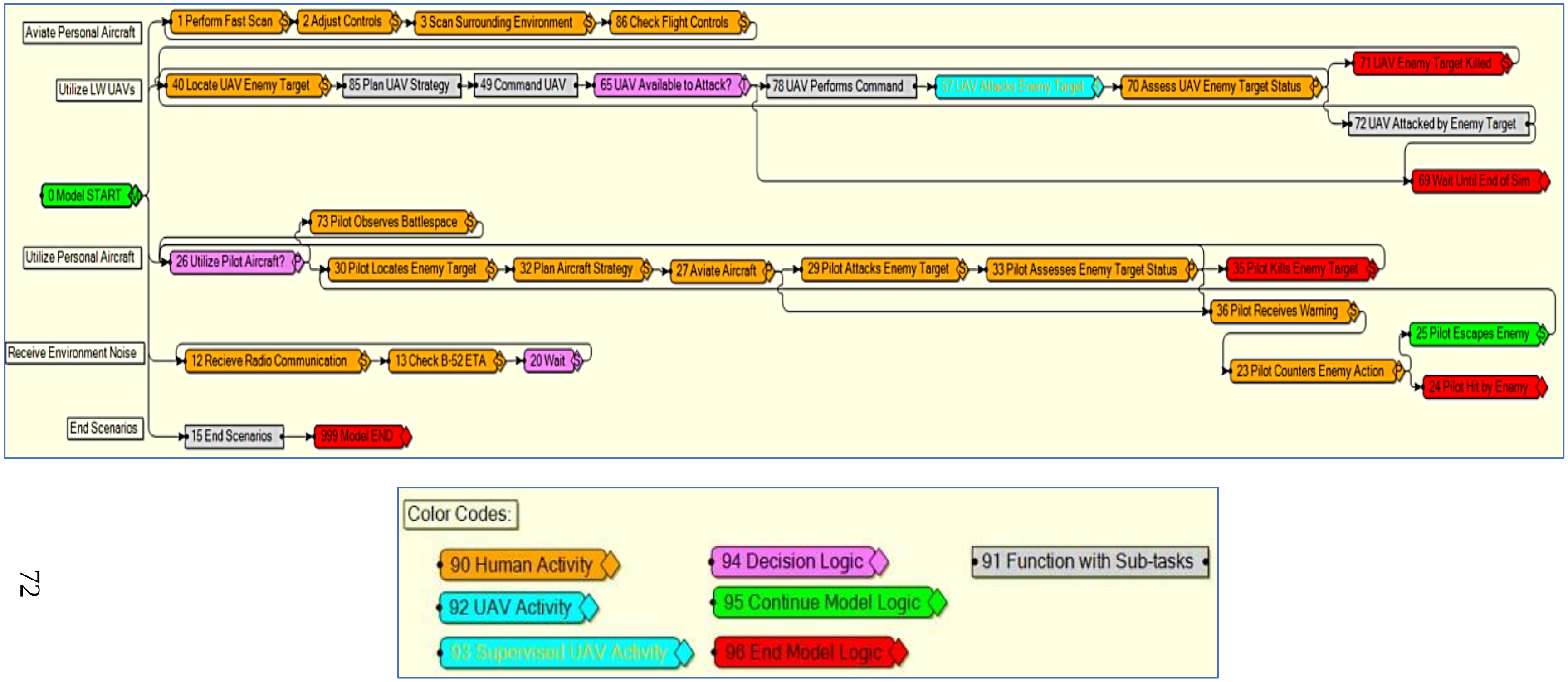


Figure 11. Baseline IMPRINT Task Network (Andrews, 2020)

With the inherent complexity of Human-Agent Interactions (HAI), this study made several assumptions in order to create a simplified IMPRINT model that could be analyzed towards the understanding of general HAI behavior. First of all, the DES assumed that all command pilots had similar levels of ability, expertise, competence, and speed. Therefore, the single model did not account for learning effects or different strategies that participants may have used. It was also assumed that all pilots utilized a "backseat" strategy to control the UAVs, meaning that the pilots forward deployed the UAVs before getting involved in the engagement themselves.

Moreover, the model focused on conditions in the peak performance region in which the human subjects arrived at their checkpoint and were actively engaged with the opponents. This meant that the segment of time in which the operators were traveling to the engagement zone was not included in the model. Furthermore, each simulation had the same conditions and did not feature any abnormal or unanticipated changes. It was assumed that any deviations in recording times did not trigger a significant decrease in model accuracy and each of the distributions applied in the model were an accurate representation of the participant pool. Finally, workload values and task times were based on data provided by a previously conducted Human-In-The-Loop (HITL) study, and as such, its applicability may be limited beyond this scope. It is noted that it may be impossible to achieve this direct comparison during an actual tactical mission.

Using this model, the effect of HAI on the pilot's cognitive workload was studied for five different conditions described in Table 7. As the amount of autonomous control abstraction increases, the number of cognitive tasks relinquished by the command pilot and reassigned to the UAVs also increases. Additionally, the amount of approval authority required before a UAV initiates an action decreases.

Table 7. IMPRINT Model Conditions for Increasing Autonomous Control Abstraction

	Condition	LHCA	Description of Pilot's Role	Description of UAV Role
1	Traditional Manned Wingman (Fully Manual)	N/A	Pilot performs all planning, decision making, and action execution for manned aircraft	No UAV involvement
2	Vector Steering (VS)	Parametric Control	Pilot flies manned aircraft; performs all planning and decision making for individual UAV movements	UAV follows specific Pilot commands
3	Pilot Directed Engagement (PDE)	Goal-Oriented Control	Pilot flies manned aircraft; performs general planning and decision making for organizational movements	UAV autonomously decides how to execute general Pilot commands
4	Tactical Battle Manager (TBM)	Goal-Oriented Control	Pilot flies manned aircraft; performs overarching planning and decision making for expected outcomes with minimal interference	UAV decides and acts autonomously, unless recommended action is vetoed by the Pilot
5	No Manned Aircraft Engagement (Combination)	Parametric and Goal-Oriented Control	Pilot flies manned aircraft; offers no assistance in attacking enemy targets, only commands UAV from afar	UAV executes pilot commands, which are a combination of VS, PDE, and TBM commands

Each level of control can be executed using a specific structure of command. For instance, the pilot may give a VS command by directing a single UAV to "turn left 45 degrees" or "fly airspeed 180 knots." Both of these commands are at a low degree of control abstraction because the UAV rapidly executes a specific, linear action in response to the pilot's command. The operator may alternatively give a PDE command by directing a group of UAVs to "form up" on a designated lead UAV and fly to "intercept target 1," meaning that the UAVs would autonomously determine how to fly in formation behind the leader and come in contact with the enemy target. This type of command is at a mid-degree of autonomous control abstraction because the UAVs autonomously decide how to orientate themselves into a position specified by the pilot. Furthermore, the pilot could give a TBM command by ordering a single UAV or formation of UAVs to "attack target 1." This means that the UAVs would use high-level intelligent reasoning to autonomously decide how to attack the target and then carry out the plan without requiring prior approval from the pilot. In each of these cases, the pilot is able to veto or intervene before the execution of a decision by an UAV. The types of commands and their corresponding conditions are summarized in Figure 12.

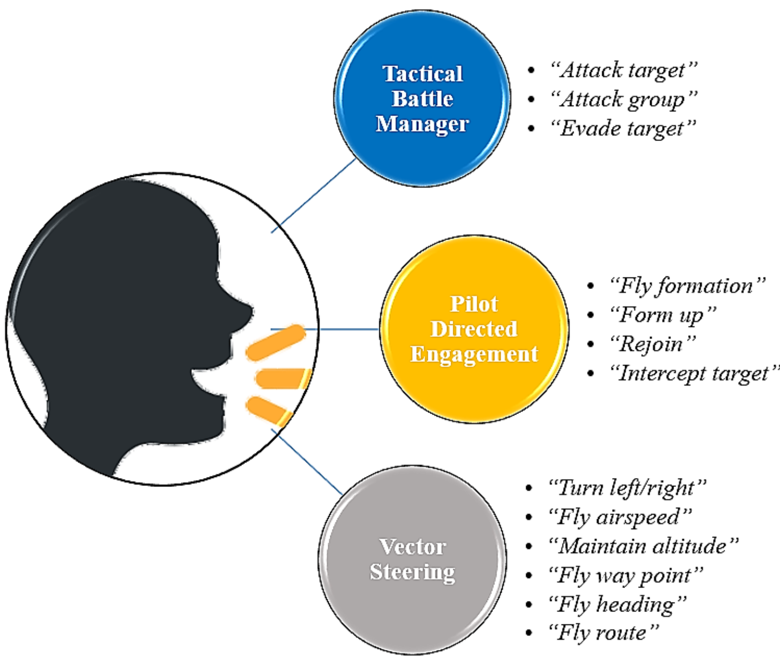


Figure 12. Examples of UAV Commands for MUM-T

For each of the five levels of UAV control described in Table 7, the baseline IMPRINT model was altered to simulate the five types of conditions. For the first condition, Traditional Manned Wingman Role, the pilot alone performed all planning, decision making, and action execution using his or her own aircraft to mirror traditional air warfare. However, the pilot is at a disadvantage as they are engaging 4 enemy aircraft without any support. Since no UAVs were deployed by the pilot, the “Utilize LW UAVs” loop was eliminated from the task network because there were no UAVs available to offer assistance to the manned aircraft. The modified baseline model is shown in Figure 13 on page 78.

For the next three conditions (VS, PDE, and TBM), the level of decision authority granted to the automation rose as the amount of autonomous control abstraction increased from VS to TBM. To simulate the increase in control abstraction, three separate conditions with varying levels of workload and task times were created, summarized in Table 8 and Table 9.

Table 8. IMPRINT Task Times and Workload Values for Commanding UAVs

Condition	**Task Time (seconds)**	**Workload Value**			
		Auditory	*Cognitive*	*Speech*	*Total*
VS	40	4.30	5.30	2.00	11.60
PDE	10	4.30	5.00	2.00	11.30
TBM	5	4.30	4.60	2.00	10.90

Table 9. IMPRINT Task Times and Workload Values for Overseeing UAVs Perform Commands

Condition	**Task Time Distribution**	**Workload Value**	
		Cognitive	*Visual*
VS	Weibull Scale: 0.39814 Shape: 1.16338	6.80	5.00
PDE	Log Logistics Scale: 3.38044 Shape: 13.74008	6.80	4.40
TBM	Weibull Scale: 114.64107 Shape: 2.77098	6.80	4.00

The modified baseline model is shown in Figure 14 on page 80. In the DES, the probability of choosing a specific task node was set to 1 to simulate one of the three conditions. For example, the probability of the pilot giving a high level command was set to 1 and the other probabilities were set to 0 to analyze the effects of only utilizing TBM commands. This method ensured that only one condition was analyzed at a time. Furthermore, the task completion time to command a UAV decreased as the level of command increased to appropriately compare the amount of time it would take each condition to execute the same action. For instance, a pilot would need to give multiple simple, low vector task commands in order to get a UAV to attack an enemy target. Conversely, the pilot would only need to give one complex, tactical task command to order the UAV to perform the same action. Finally, the task probabilities and time distributions for a UAV to execute a command were modeled after the data collected from the ATACM experiments (see Appendix 3).

For the final condition, No Manned Aircraft Engagement, the pilot offered no assistance to the UAVs when attacking an enemy target. Instead, the UAVs were forward deployed and commanded by the pilot using either VS, PDE, or TBM commands to attack the adversaries. Accordingly, the "Utilize Personal Aircraft" loop was eliminated from the task network since the pilot was not engaging any of the enemy targets personally, shown in Figure 15 on page 80. This scenario was built to evaluate whether or not the involvement of a pilot was worth the risks associated with participating in the engagement themselves.

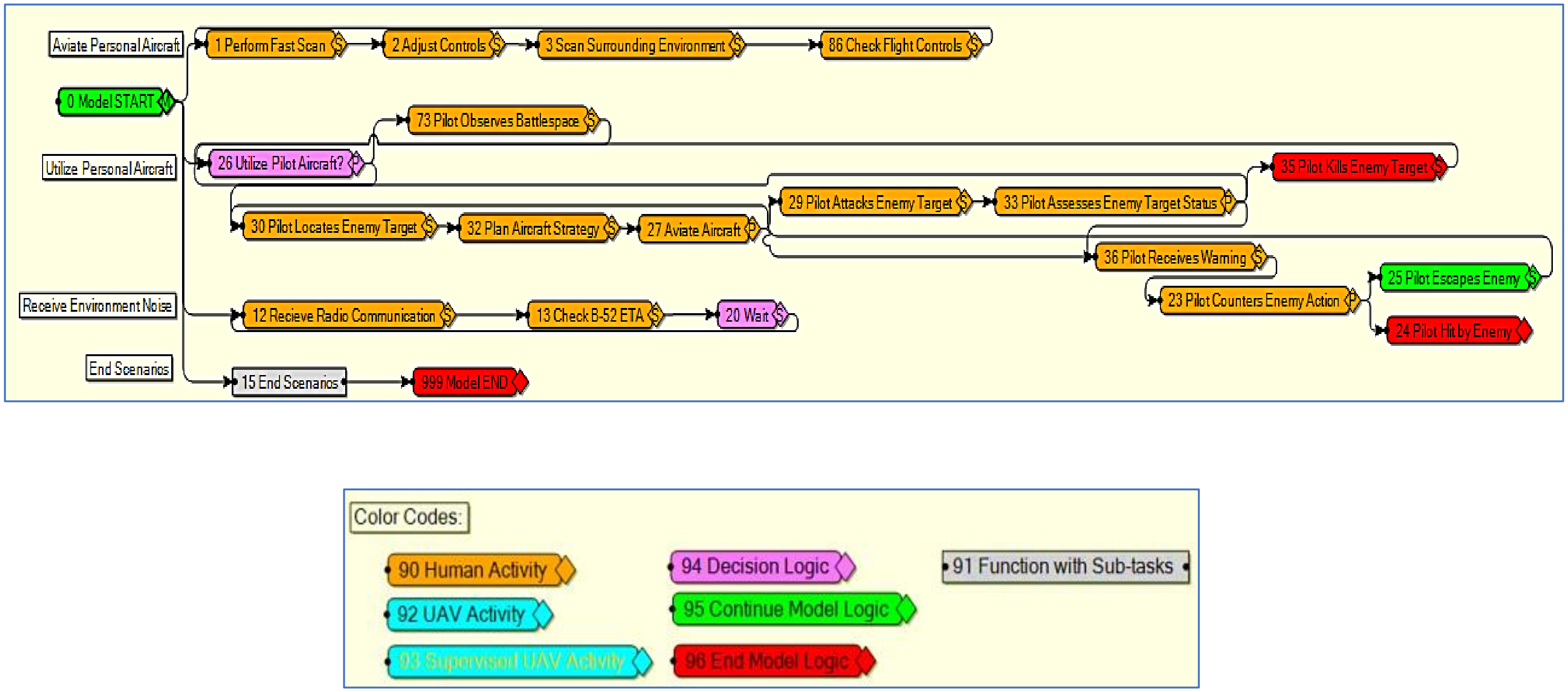


Figure 13. Traditional Manned Wingman Role IMPRINT Task Network

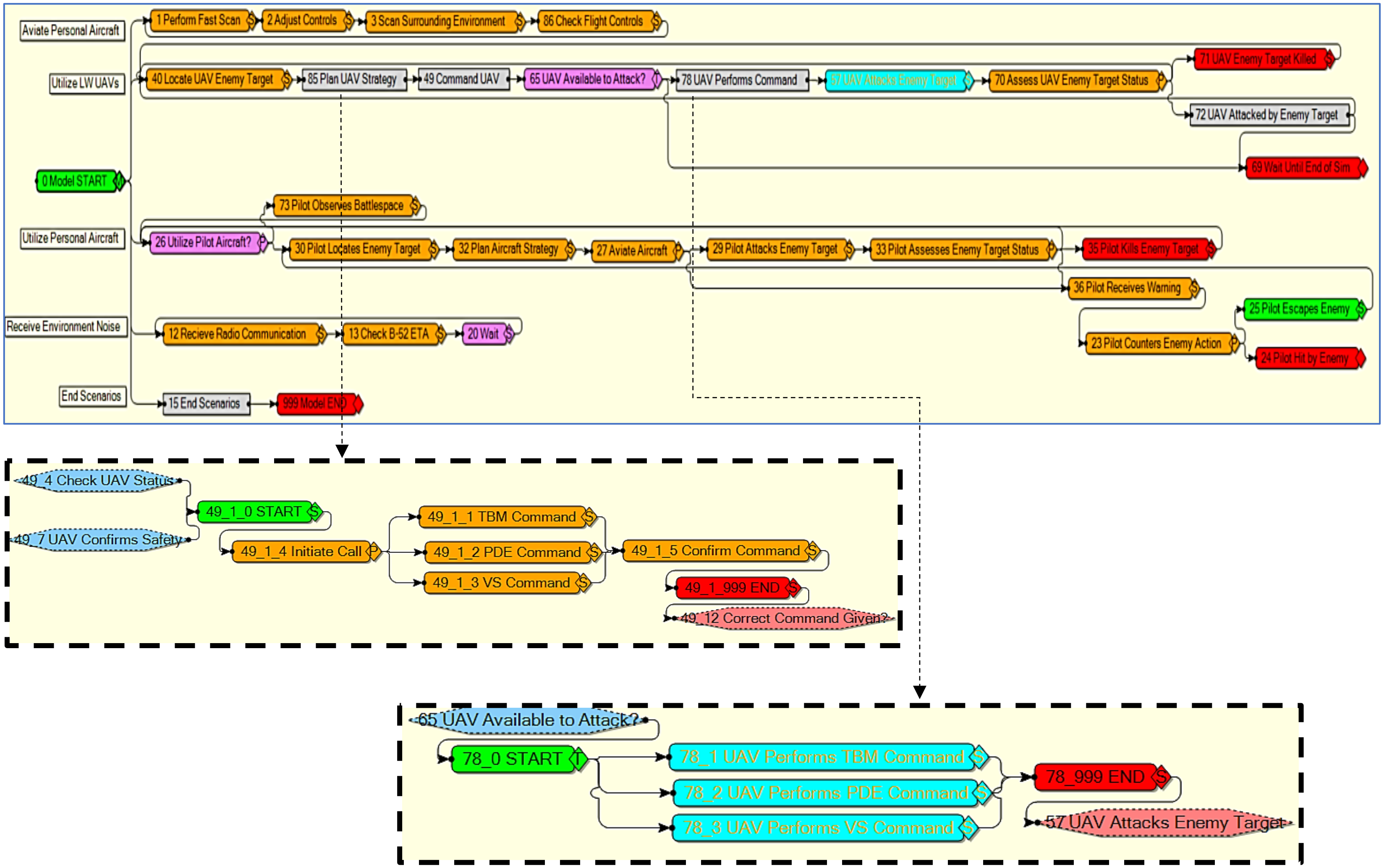
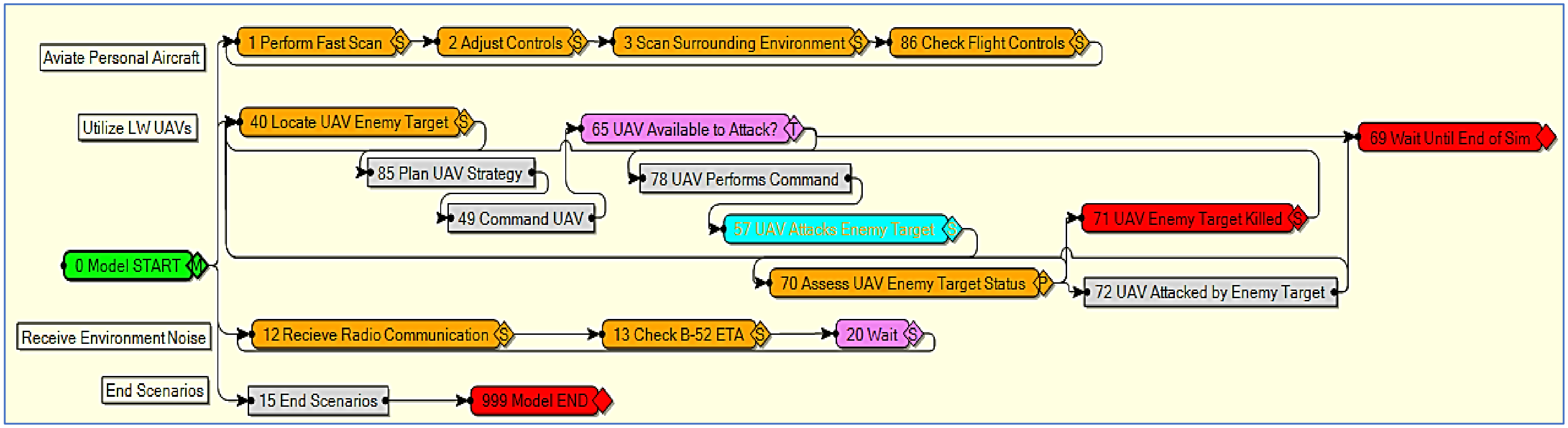


Figure 14. IMPRINT Task Network for VS/PDE/TBM Commands



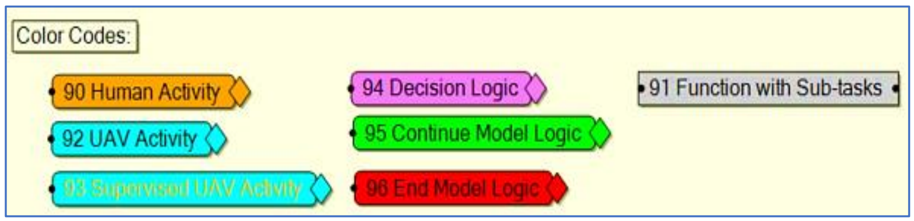


Figure 15. No Manned Aircraft Engagement IMPRINT Task Network

Analysis and Results

In this DES, three enemy targets fought against one pilot and three UAVs. The independent variable was the amount of autonomous control abstraction given by the pilot to the UAVs. The dependent variables were the mission performance and mental workload of the pilot during a simulation run. One thousand trials were run in IMPRINT for each of the five conditions listed in Table 7. Mission performance was measured by calculating the number of enemy targets that survived compared to the number of UAVs that survived. In addition, the workload of the pilot was determined using the workload profiles and VACP scores gathered from 30 trials, producing five different time-persistent averages for each model.

Mission Performance Analysis

The total number of UAVs remaining after the simulated engagement are shown in Figure 16, and the total number of enemy targets are shown in Figure 17. The number of UAVs remaining was not calculated for the Traditional Manned Wingman, fully manual model as no UAVs were utilized in this condition.

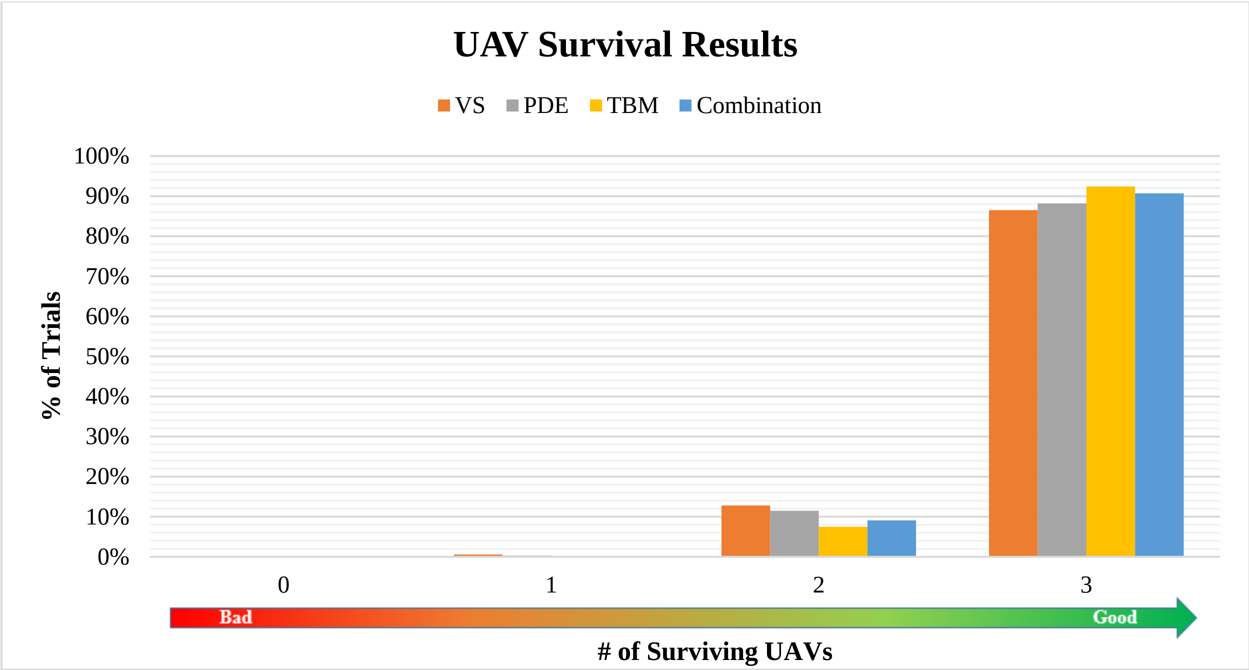


Figure 16. Graph of UAV Survival Results for Conditions 2-5

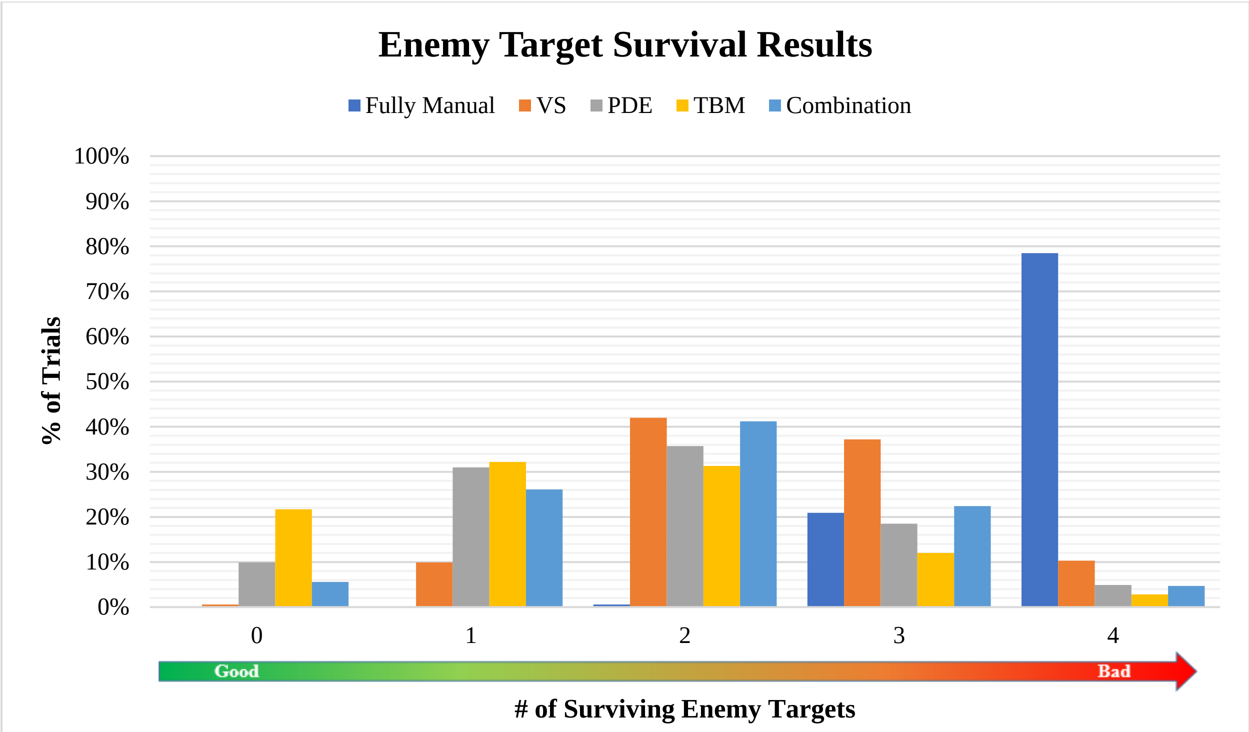


Figure 17. Graph of UAV Survival Results for Conditions 1-5

Using the UAV survival results obtained from the DES, a statically significant difference was observed between the means of each condition as determined by a one-way Analysis of Variance (ANOVA) using a 95% confidence interval (see Appendix 4). This result provided statistical evidence that increasing the level of autonomous control abstraction did effect the UAV survival rate. However, it was observed that there was no statically significant difference among sample means for VS-PDE, PDE-Combination, and TBM-Combination pairs as determined by a Tukey Honestly Significant Difference (HSD) test (see Appendix 4).

The percentage of surviving UAVs per 1,000 trials (see Table 10) and average number of surviving UAVs per trial (see Table 11) were then calculated to decide the amount of autonomous control abstraction that resulted in the greatest number of UAVs that survived for the most number of trials. According to the calculations, employing TBM commands or a combination of VS, PDE, and TBM commands resulted in the greatest number of UAVs surviving over 1,000 trials. Whereas employing VS resulted in the least number of UAVs surviving over 1,000 trials. This result was expected because the UAVs that had greater autonomy were able to make rapid decisions and act swiftly, since they did not need to wait for pilot input to evade from enemy fire.

Table 10. Percentage of Surviving UAVs per 1,000 Trials

# of Surviving UAVs	**% of Trials Resulting in Surviving UAVs**			
	VS	*PDE*	*TBM*	*Combination*
0	0%	0%	0%	0%
1	1%	0%	0%	0%
2	13%	12%	8%	9%
3	87%	88%	92%	91%

Table 11. Average Number of Surviving UAVs per Trial

Condition	Average # of Surviving UAVs
VS	2.857
PDE	2.879
TBM	2.922
Combination	2.905

Using the enemy target survival results obtained from the DES, a statically significant difference was also observed between the means of each condition as determined by a one-way Analysis of Variance (ANOVA) using a 95% confidence interval (see Appendix 4). Furthermore, it was observed that all of the pairs were statistically different from each other as determined by a Tukey HSD test (see Appendix 4). This result provided statistical evidence that increasing the amount of autonomous control abstraction did effect the enemy target survival rate.

The percentage of surviving enemy targets per 1,000 trials (see Table 12) and average number of surviving enemy targets per trial (see Table 13) were then calculated to determine how much of the operator's cognitive tasks should be relinquished by the command pilot and reassigned to the UAVs to result in the most enemy targets destroyed for the greatest number of trials. According to the calculations, deploying TBM commands resulted in the greatest number of enemy targets getting killed over 1,000 trials. The Traditional Manned Wingman role, which required no automation, resulted in the least number of enemy targets getting killed over 1,000 trials. This result was expected assuming that the automation is nearly as effective as the pilot at commanding the UAVS. Under this assumption, using TBM commands would result in the least number of enemy targets surviving because the pilot can quickly command multiple UAVs to perform a high-level action using a single verbal command. On the other hand, the pilot would have to exert more time and effort to attack the enemy target with his or her manned aircraft or commanding the UAVs using lower amounts of autonomous control abstraction.

Table 12. Percentage of Surviving Enemy Targets per 1,000 Trials

# of Surviving Enemies	**% of Trials Resulting in Surviving Enemies**				
	Fully Manual	*VS*	*PDE*	*TBM*	*Combination*
0	0%	1%	10%	22%	6%
1	0%	10%	31%	32%	26%
2	1%	42%	36%	31%	41%
3	21%	37%	19%	12%	22%
4	79%	10%	5%	3%	5%

Table 13. Average Number of Surviving Enemy Targets per Trial

Condition	**Average # of Surviving Enemy Targets**
Fully Manual	3.779
VS	2.467
PDE	1.775
TBM	1.420
Combination	1.945

According to both the UAVs and enemy target survival results, increasing the amount of autonomous control abstraction improved the mission performance of the human-agent system. Utilizing TBM commands or a combination of VS, PDE, and TBM commands improved the survival of the UAVs and utilizing just TBM commands increased the likelihood of killing enemy targets. Thus, the integration of MUM-T through TBM produced the highest level of mission performance in this task scenario.

Workload Profile Analysis

In addition to the mission performance, the pilot's cognitive workload levels were analyzed for the five model conditions described in Table 7. The total objective workload experienced by the operator at each instant of a single simulation run was calculated by IMPRINT and graphed in Figure 18. Events that were above a workload level of 60 were considered to be near or above the saturation threshold where the system imposed more work than the pilot could effectively perform (Mitchell, 2003; Schneider & McGrogan, 2011). In an ideal mission scenario, all workload levels would be below 60.

The workload profile shown in Figure 18 illustrated the amount of mental effort required by the pilot to command three UAVs using varying amounts of autonomous control abstraction. The graph provided insight into how the pilot's cognitive workload levels were affected by changing how much of the pilot's cognitive tasks should be relinquished by the operator and reassigned to the UAVs in a specific trial run.

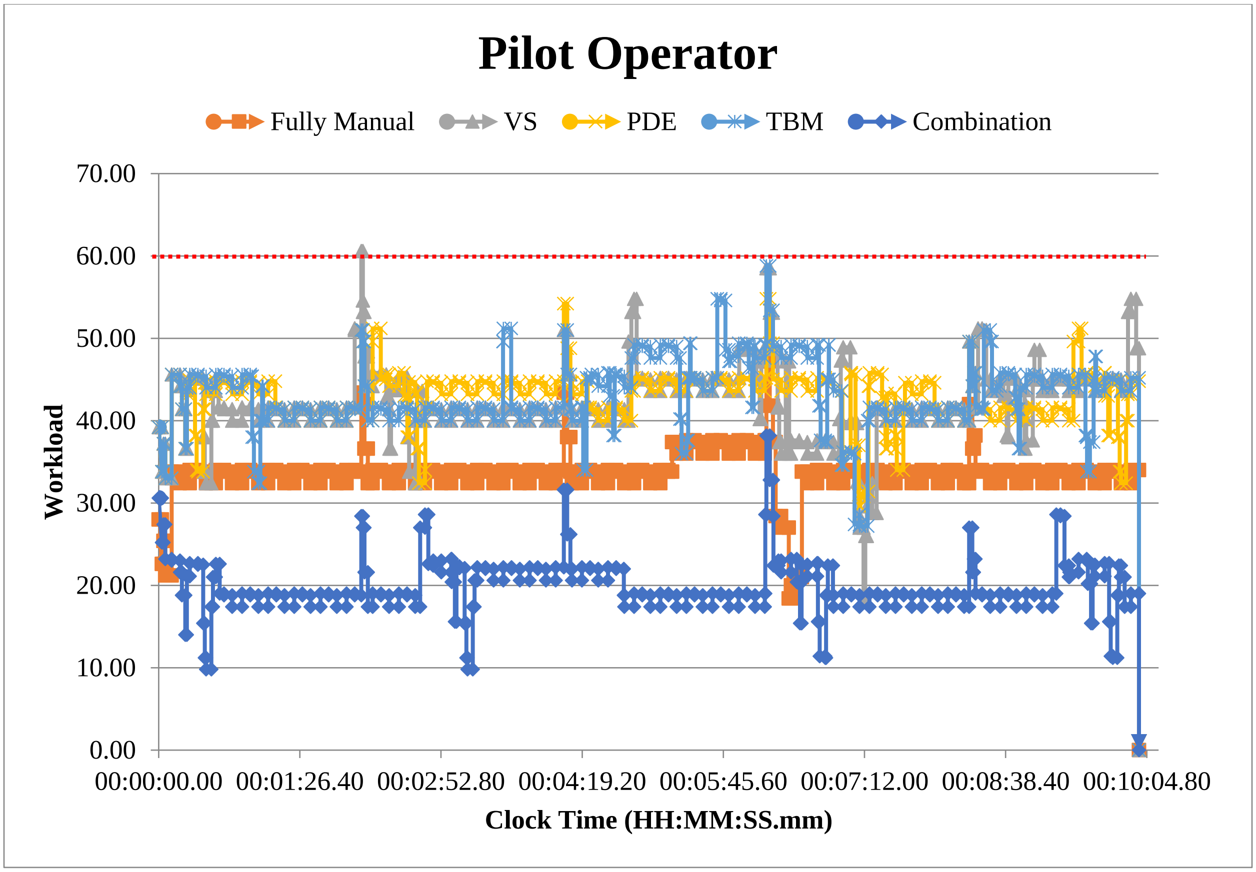


Figure 18. IMPRINT Workload Profile for Pilot in Model Conditions 1-5

According to Figure 18, the pilot experienced the highest levels of cognitive workload when utilizing VS to command the three UAVs. The workload saturation level was surpassed in this condition, indicating the pilot's inability to effectively or safely operate both manned and unmanned aircraft at the same time. The pilot experienced the next two highest levels of workload when employing PDE and then TBM commands to control the UAVs. The next lowest levels of workload were experienced in the fully manual, Traditional Manned Wingman role, which was anticipated since the pilot only utilized the manned aircraft to attack enemy targets. Finally, the pilot experienced the lowest levels of workload in the combination, No Manned Aircraft Engagement role, which transitioned the Pilot to a supervising role and transferred the burden of fighting the enemy targets to the UAVs. The results indicated a large drop in workload levels when the UAVs were forward deployed using a combination of VS, PDE, and TBM commands.

Time-Persistent Average Workload Analysis

Using the VACP workload values from IMPRINT, a single representative workload value was also computed by calculating the time-persistent average across 30 DES for the first four model conditions. The time-persistent average illustrated how hard the pilot worked as a whole to command the three UAVs. According to Table 14, the pilot experienced the lowest time-persistent average workload of 19.77 when using the No Manned Aircraft Engagement role and the highest time-persistent average workload of 43.22 when only using VS commands. The results indicated that the pilot's cognitive workload for a large portion of the time was below the saturation level for each model condition, but it varied significantly throughout the simulation. Furthermore, the pilot experienced increased levels of workload when the amount of autonomous control abstraction decreased. This finding is consistent with the results obtained from the analysis of the workload profile.

Table 14. Time-Persistent Average of the IMPRINT Workload Profile for Conditions 1-5

Condition	**Pilot Operator**		
	Minimum	**Maximum**	**Time Persistent Average**
Fully Manual	16.93	46.27	33.83
VS	24.61	56.57	43.22
PDE	23.36	56.72	42.73
TBM	24.46	56.29	42.15
Combination	9.80	32.66	19.77

Although the burden of operator management decreased as autonomy increased, increasing autonomy does not always improve the overall performance of the human-agent system. According to research conducted by Johnson et al. (2012), a decrease in mental workload levels does not necessarily equate to increased effectiveness. Therefore, both factors must be considered to appropriately determine what level of command the UAVs should be automated to reduce operator workload and increase mission performance. Through an analysis of the mission performance, workload profile, and time-persistent averages, it was determined that using a combination of VS, PDE, and TBM commands would lead to increased performance for the human-agent team. The incorporation of all three commands would ensure that the pilot is able to control both the manned and unmanned aircraft, while having enough control over the UAVs to anticipate their behavior. Furthermore, the forward deployment of the UAVs permits the pilots to distance themselves from enemy fire, thus increasing their chances of survival in air-to-air warfare.

Conclusion

The research performed in this study sought to use DES to explore how changes in autonomy affected the human-agent team's mission performance and the pilot's cognitive workload. This was accomplished by building an IMPRINT model to investigate the level of control abstraction the UAVs should be automated to reduce operator workload and increase mission performance in the flight operation task. Although a reduction in human workload is both the common expectation and the major motivation for automation (M. Johnson et al., 2012), system designers for should not automatically increase the autonomy of the UAVs without addressing the operator's ability to understand what is happening and anticipate the agent's behavior. For this reason, the UAVs should be automated to handle a varying amount of autonomous control abstraction using a combination of VS, PDE, and TBM commands to achieve increased mission performance and maintain the pilot's cognitive workload at a manageable level.

For future development, attention should be devoted to determine how many UAVs a single pilot can effectively operate simultaneously. By further studying the impact of MUM-T on mission effectiveness and its effect on the pilots who will be commanding them, the U.S. Air Force will be one step closer to successfully incorporating MUM-T into flight operations. Thus, changing the way that the aviation community has thought about piloting for over 100 years.

V. Conclusion and Recommendations

Chapter Overview

The purpose of this chapter is to answer the investigative questions, provide insights into the significance and limitations of the research, recommend a course of action, and propose future research. A novel Discrete Event Simulation (DES) was developed in this research to evaluate the potential effects of Manned-Unmanned Teaming (MUM-T) on the pilot's cognitive workload and overall mission performance. The results of this research provided insights into the potential benefits or issues that may arise from incorporating MUM-T into air operations. It also revealed the amount of autonomous control abstraction that have the largest impact in reducing operator workload and increasing mission performance to provide Human-Agent Interactions (HAI) recommendations for system improvements.

Answers to Research Questions

The following research questions were addressed to fully answer the overarching inquiry of how HAI benefits or degrades pilot workload and mission performance:

1. ***How does the use of MUM-T affect the pilot's cognitive workload during combat mission events?***

   The results of the simulation experiments indicated that the command pilot generally experienced a manageable level of workload when commanding 3 UAVs against 4 enemy targets using a vocally commanded interface. However, peaks in workload occurred for the pilot during periods of high communications load and this communication may be degraded or delayed during air-to-air engagements. This is an area of concern for system designers because it may be difficult for pilots to balance radio calls while commanding UAVs under normal operating conditions or high G-stress.

2. ***How does the use of MUM-T affect the human-agent team's mission performance during combat mission events?***

   It was concluded that the mission performance was significantly improved by the use of 3 UAVs against 4 enemy targets. According to the DES results, the human-agent team was 18.40% more successful on average in striking all four enemy targets than the manned-only condition.

3. ***To what degree of autonomous control abstraction should UAVs perform at to reduce operator workload in a flight operation task?***

   The results obtained from the alternative simulation experiments revealed the largest drop in workload levels when the UAVs were forward deployed using a combination of Vector Steering, Pilot Directed Engagement, and Tactical Battle Manager commands. Therefore, the UAVs should be automated to handle varying levels of autonomous control abstraction to maintain the pilot's cognitive workload at a manageable workload level when commanding 3 UAVs against 4 enemy targets in an air-to-air operation.

4. ***To what degree of autonomous control abstraction should UAVs perform at to increase mission performance in a flight operation task?***

   According to DES results, utilizing either Tactical Battle Manager commands or a combination of Vector Steering, Pilot Directed Engagement, and Tactical Battle Manager commands improved the survivability of the UAVs. However, only the use of Tactical Battle Manager commands produced the highest likelihood of killing all 4 of the enemy targets. Therefore, it was concluded that the integration of MUM-T through Tactical Battle Management, a high degree of autonomous control abstraction, would enable the human-agent team to achieve increased mission performance in terms of successful adversary strikes as well as UAV and pilot survivability.

Assumptions and Limitations

Creating an IMPRINT model required task analyses, direct observations, and data collection of a system. However, MUM-T had yet to be deployed in an operational environment. Consequently, this research was reliant on information provided by Subject Matter Experts (SMEs) and data collected from a Human-In-The-Loop (HITL) study performed by the 711 Human Performance Wing (HPW) at Wright Patterson Air Force Base.

While the pilots were non-experts within a virtual environment, it was assumed that the human participants and tasks were sufficiently representative of MUM-T operators and operations to effectively evaluate performance and workload impacts of automation. It was also assumed that the human subjects involved in the Autonomy for Air Combat Missions (ATACM) study gave their maximum effort and were trained to a stable skill level prior to data collection, minimizing any learning effects across the trials. Furthermore, it was assumed that the randomized order of the conditions resulted in no order effects and did not affect the workload or physiological changes in this investigation. Finally, the SMEs estimates were assumed to be accurate approximations to real-world data, which was justified because the SMEs had experience developing and using the ATACM environment.

With the inherent complexity of HAI, this study makes several assumptions in order to create a simplified IMPRINT model that can be analyzed towards the understanding of general HAI behavior. First of all, the DES assumed that all command pilots have similar levels of ability, expertise, competence, and speed. Therefore, the single model did not account for learning effects or different strategies that participants may have used. It was also assumed that all pilots utilized a “backseat” strategy to command the Unmanned Aerial Vehicles (UAV), meaning that the pilots forward deployed the UAVs before getting involved in the engagement themselves.

Moreover, the model focused on conditions in the peak performance region in which the human subjects arrived at their checkpoint and were actively engaged with the opponents. This meant that the segment of time in which the operators were traveling to the engagement zone was not included in the model. Furthermore, each simulation had the same conditions and did not feature any abnormal or unanticipated changes. It was also assumed that any deviations in recording times did not trigger a significant decrease in model accuracy and each of the distributions applied in the model were an accurate representation of the participant pool. Finally, workload values and task times were based on data provided by the 711 HPW, and as such, its applicability may be limited beyond this scope. It is noted that it may be impossible to achieve this direct comparison during an actual tactical mission.

Recommendation for Actions

The recommended action is to develop UAVs that are capable of handling a combination of Vector Steering, Pilot Directed, and Tactical Battle Manager commands. Although the burden of operator management decreased as autonomy increased, increasing autonomy does not always improve the overall performance of the human-agent system. According to research conducted by Johnson et al. (2012), a decrease in mental workload levels does not necessarily equate to increased effectiveness. Therefore, both factors must be considered to appropriately determine what level of command UAVs should be automated to reduce operator workload and increase mission performance. Through an analysis of the mission performance, workload profile, and time-persistent averages, it was determined that using a combination of Vector Steering, Pilot

Directed, and Tactical Battle Manager commands would lead to increased performance for the human-agent team. The incorporation of all three commands would ensure that the pilot is able to control both the manned and unmanned aircraft, while having enough control over the UAVs to anticipate their behavior. In addition, the forward deployment of the UAVs permits the pilots to distance themselves from enemy fire, thus increasing their chances of survival in air-to-air air warfare.

Furthermore, system designers should be cognizant of the potential for pilots to experience peaks in workload levels when commanding 3 UAVs against 4 enemy targets. The command pilot bears the weight of the combat effort and will need to deploy capabilities from the UAVs in addition to controlling the manned aircraft. The challenge of maintaining close control of the UAVs could be difficult for pilots to maintain during periods of high communications load, which could lead to a degrade or delay in communication capabilities during air-to-air engagements. Therefore, system designers should design a pilot-vehicle interface that is conducive to maintaining a manageable level of workload between the pilot in the cockpit and the UAVs in the sky.

Recommendation for Future Research

For future development, the DES should be updated to examine additional alternative scenarios. While these results provided insight into using different automation controls for MUM-T operations, the presented research was limited to data provided by the ATACM experiment. The next step would be to gather data that exists outside of a HITL experiment in order to develop a model that more realistically captures HAI between pilots and their UAVs in an operational environment. Once this type of data becomes available, an improved model could be used to determine how many UAVs a single pilot can effectively operate simultaneously and in what type of formation are they best commanded. The improved model would further examine the relationship between stages and levels to discern which combinations work together optimally to better capture human-agent system behavior. This information could enable system designers to test and evaluate multiple configurations of MUM-T systems in a short period of time and at a marginal cost.

When making automation implementation tradeoffs, other factors, such as situation awareness, reliability, and trust may also impact operator workload and system performance. Future work should seek to identify these factors and examine their impacts with on the pilot's cognitive workload and the mission performance with regards to the different combinations of human-agent teaming. If one combination has less sensitivity than another, it may be prudent to choose the less sensitive combination.

In addition, future research should develop a new autonomous control taxonomy that more appropriately describes the relationship between humans and agents in MUM-T. Although there has been some development in this area of research with the five Levels of Human Control Abstraction (LHCA) (C. D. Johnson et al., 2017) and the Interdependence Analysis Tool (IAT) (M. Johnson, Vignati, et al., 2018), progress still needs to be made to combine these approaches to provide a more comprehensive model that fully characterizes the division of work and interdependencies between the human and the agent.

In the case of MUM-T, there were some discrepancies between the LHCA and the degrees of automation for MUM-T. As the LHCA increased from Parametric Control to Goal-Oriented Control, the pilot's level of responsibility decreased and the automation's level of responsibility increased. However, this was not a binary relationship where the human operator completely relinquished all safety and regulation responsibilities to the automation. For instance, consistent with the LHCA framework, the UAVs had more capabilities and responsibilities when issued a PDE command than a VS command. However, the automation did not completely relieve the pilot of safety monitoring and obstacle avoidance, as is described Goal-Oriented Control. The pilot was still expected to perform this duty and intervene to prevent the loss of an UAV from enemy fire.

Furthermore, there was not enough precision to fully capture the nuances between the continuum of human responsibilities and degrees of automation for MUM-T. For example, there was a difference between the pilot giving a PDE or a TBM command. According to the IAT, the pilot would have fewer perception and cognition responsibilities when giving a TBM command in comparison to a PDE command. Yet, both commands were categorized under Goal-Oriented Control according to LHCA. Therefore, the LHCA frameworks needs further refinement to distinguish different control approaches with an LHCA level. It is conceivable that design tradeoffs frequently occur within LHCA levels rather than between levels. A stronger model could be developed by leveraging and combining the strengths and features of LHCA and IAT to help designers better assess the potential interdependencies between workload and workflows for the human and the agent in MUM-T systems.

Summary

The findings presented in this research are a significant step towards simulating the complexities of real-world activities by mirroring the highly dynamic nature of realistic military operations in a virtual environment. MUM-T had never been modeled using IMPRINT before this research was conducted. Not only did this study develop an original DES, but it also provided insights into the effects of MUM-T on the pilot’s cognitive workload levels and the human-agent team’s overall mission performance. Using this information, system designers from the 711 HPW can integrate the results obtained from this study into future human-agent system design considerations. By studying the impact of MUM-T on mission performance and its effect on the pilots who will be commanding them, the U.S. Air Force will be one step closer to successfully incorporating MUM-T into flight operations. Thus, changing the way that the aviation community has thought about piloting for over 100 years.

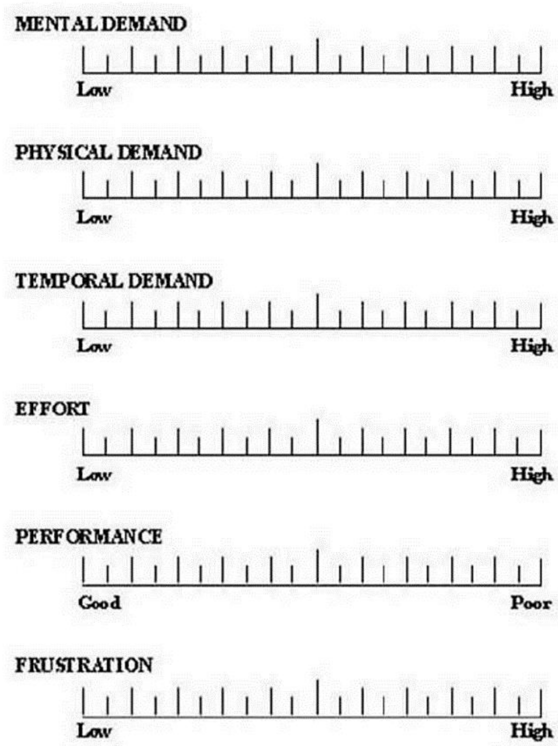
Appendices

Appendix 1: NASA-TLX Workload Rating Scale

Table 15 describes the standardized NASA-TLX workload surveys administered to ATACM study subject participants (Hart & Staveland, 1988).

Table 15. NASA-TLX Workload Rating Sale

Category	End Points	Questions
Mental Demand	Low/High	How much mental and perceptual activity was required (i.e. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (i.e. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Perceived Performance	Low/High	How successful were you in performing the goals of task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?



Appendix 2: VACP Workload Rating Scale

Table 16 describes the standardized VACP values used in IMPRINT (Alion Science and Technology Corporation, 2009). The scale was derived from (Bierbaum et al., 1989):

Table 16. VACP Channel Workload Rating Scale

Value	Descriptors
	VISUAL
0.0	No Visual Activity
1.0	Visually Register/Detect (detect occurrence of image)
3.0	Visually Inspect/Check (discrete inspection/static condition)
4.0	Visually Locate/Align (selective orientation)
4.4	Visually Track/Follow (maintain orientation)
5.0	Visually Discriminate (detect visual difference)
5.1	Visually Read (symbol)
6.0	Visually Scan/Search/Monitor (continuous/serial inspection, multiple conditions)
	AUDITORY
0.0	No Auditory Activity
1.0	Detect/Register Sound (detect occurrence of sound).
2.0	Orient to Sound (general orientation/attention)
3.0	Interpret Semantic Content (speech, simple, 1-2 words)
4.2	Orient to Sound (selective orientation/attention)
4.3	Verify Auditory Feedback (detect occurrence of anticipated sound)
6.0	Interpret Semantic Content (speech, complex, sentence)
6.6	Discriminate Sound Characteristics (detect auditory differences)
7.0	Interpret Sound Patterns (pulse rates, etc.)
	COGNITIVE
0.0	No Cognitive Activity
1.0	Automatic (simple association)
1.2	Alternative Selection
4.6	Evaluation/Judgment (consider single aspect)
5.0	Sign/Signal Recognition
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion

	<u>FINE MOTOR</u>
0.0	No Fine Motor Activity

2.2	Discrete Actuation (button, toggle, trigger)
2.6	Continuous Adjustive (flight controls, sensor control)
4.6	Manipulative (tracking)
5.5	Discrete Adjustment (rotary, vertical thumbwheel, lever position)
6.5	Symbolic Production (writing)
7.0	Serial Discrete Manipulation (keyboard entries)
	<u>GROSS MOTOR</u>
0.0	No Gross Motor Activity
1.0	Walking on level terrain
2.0	Walking on uneven terrain
3.0	Jogging on level terrain
3.5	Heavy lifting
5.0	Jogging on uneven terrain
6.0	Complex climbing
	<u>SPEECH</u>
0.0	No speech activity
2.0	Simple (1-2 words)
4.0	Complex (Sentence)
	<u>TACTILE - feeling feedback</u>
0.0	No tactile activity
1.0	Alerting
2.0	Simple discrimination
4.0	Complex symbolic information

Appendix 3: IMPRINT Baseline Model Task Network Development & Validation

Phase 1: Conceptual Model

The first step in developing a usable baseline simulation model was to formulate a conceptual model of the human-agent system in order to ensure that all tasks, resources, and process flows were accurately captured. To develop this framework, SMEs from the ATACM study provided a general description of the activities involved in performing a given scenario, illustrated in Figure 19 and Figure 20 on pages 102-103. The activity diagrams help illustrate the type of activities participants completed throughout the ATACM trials.

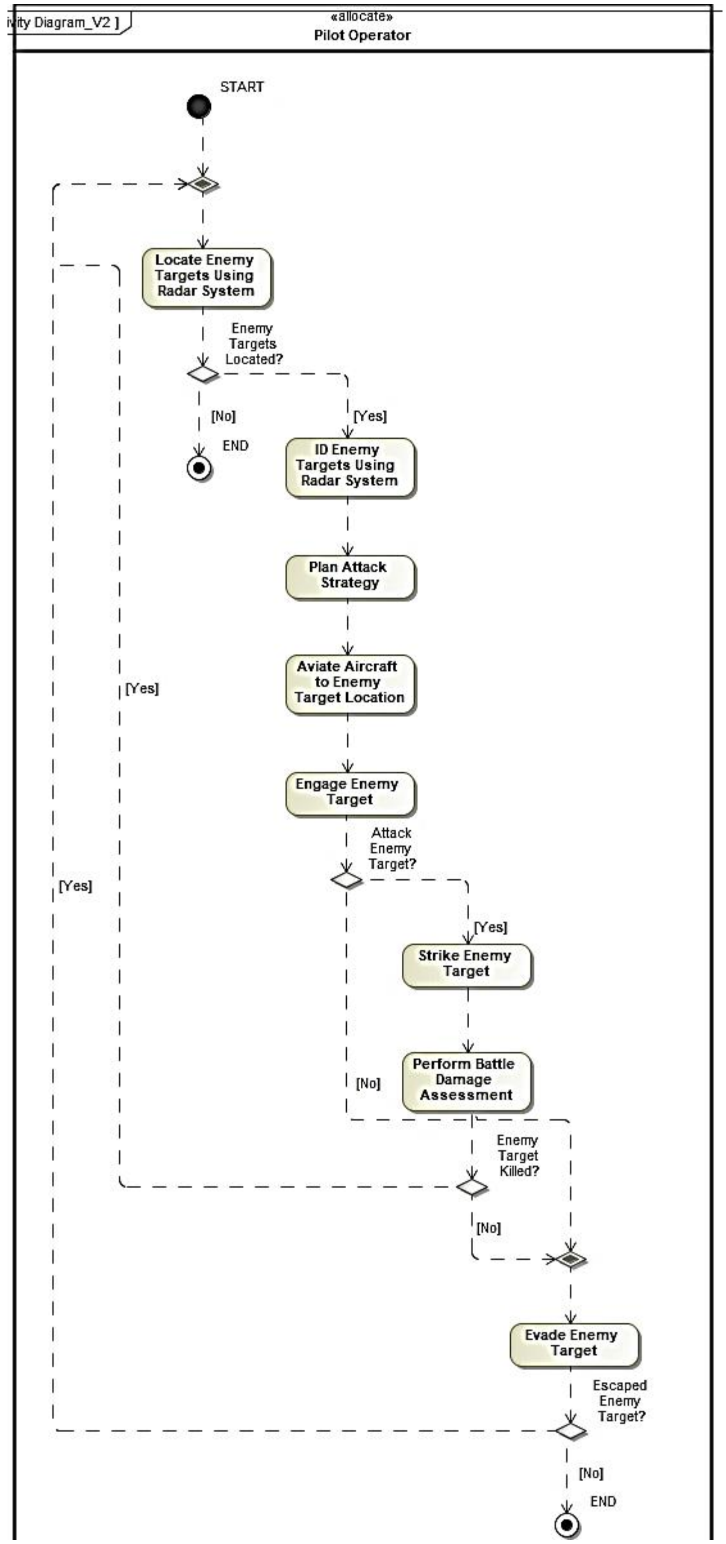


Figure 19. Activity Diagram Illustrating Pilot Utilizing Personal Aircraft

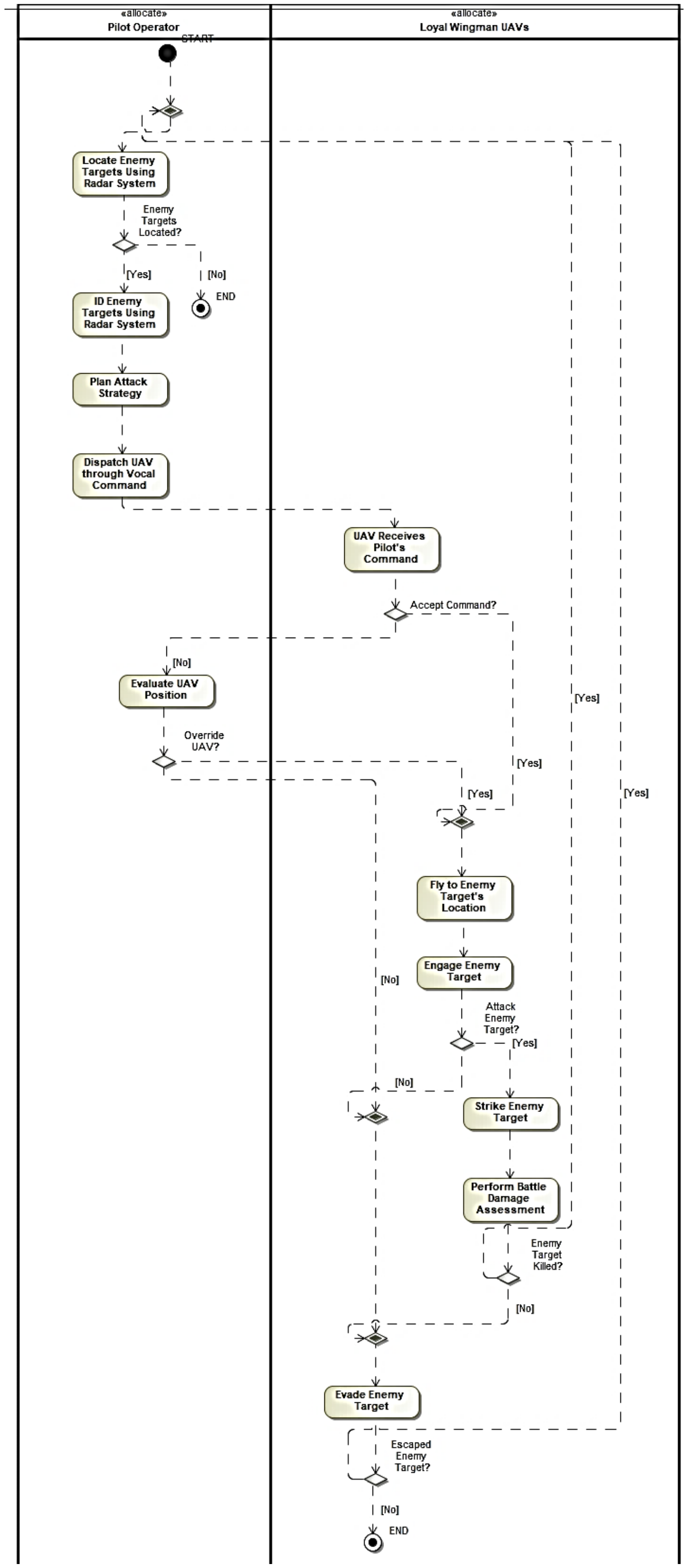


Figure 20. Activity Diagram Illustrating Pilot Utilizing UAVs

Phase 2: Task Analysis

The task networks developed in Figure 19 and Figure 20 set the foundation for the task network later developed in IMPRINT. Using IMPRINT, the flow of actions and decision logic captured in the activity diagrams were transferred to the DES environment. As shown in Figure 21 on the next page, the baseline task network model was composed of four different task loops and one logic loop: 1) Aviate Personal Aircraft, 2) Utilize UAVs, 3) Utilize Personal Aircraft, 4) Receive Environment Noise, and 5) End Scenarios.

The first task loop, “Aviate Personal Aircraft,” included basic tasks such as adjusting the flight controls or scanning the surrounding environment that the pilot performed when operating his or her own aircraft. The second task loop, “Utilize LW UAVs,” included tasks such as commanding the UAV or supervising UAV attacks, which the pilot executed to deploy the UAVs. The third task loop, “Utilize Personal Aircraft,” included tasks such as aviating the manned aircraft or attacking the adversary target, which the pilot performed in order to utilize his or her own aircraft to attack the enemy. The fourth task loop, “Receive Environment Noise,” included the workload associated with receiving audio notifications over the radio. All four of these task loops ran in parallel with one another as it was assumed that the pilot performed these activities concurrently. The final logic loop, “End Scenarios,” included tasks that would trigger the DES to end if any of the stopping scenarios were fulfilled. The logic loop also ran concurrently with the other task loops in order for the software to evaluate whether or not the simulation satisfied one of the ending conditions.

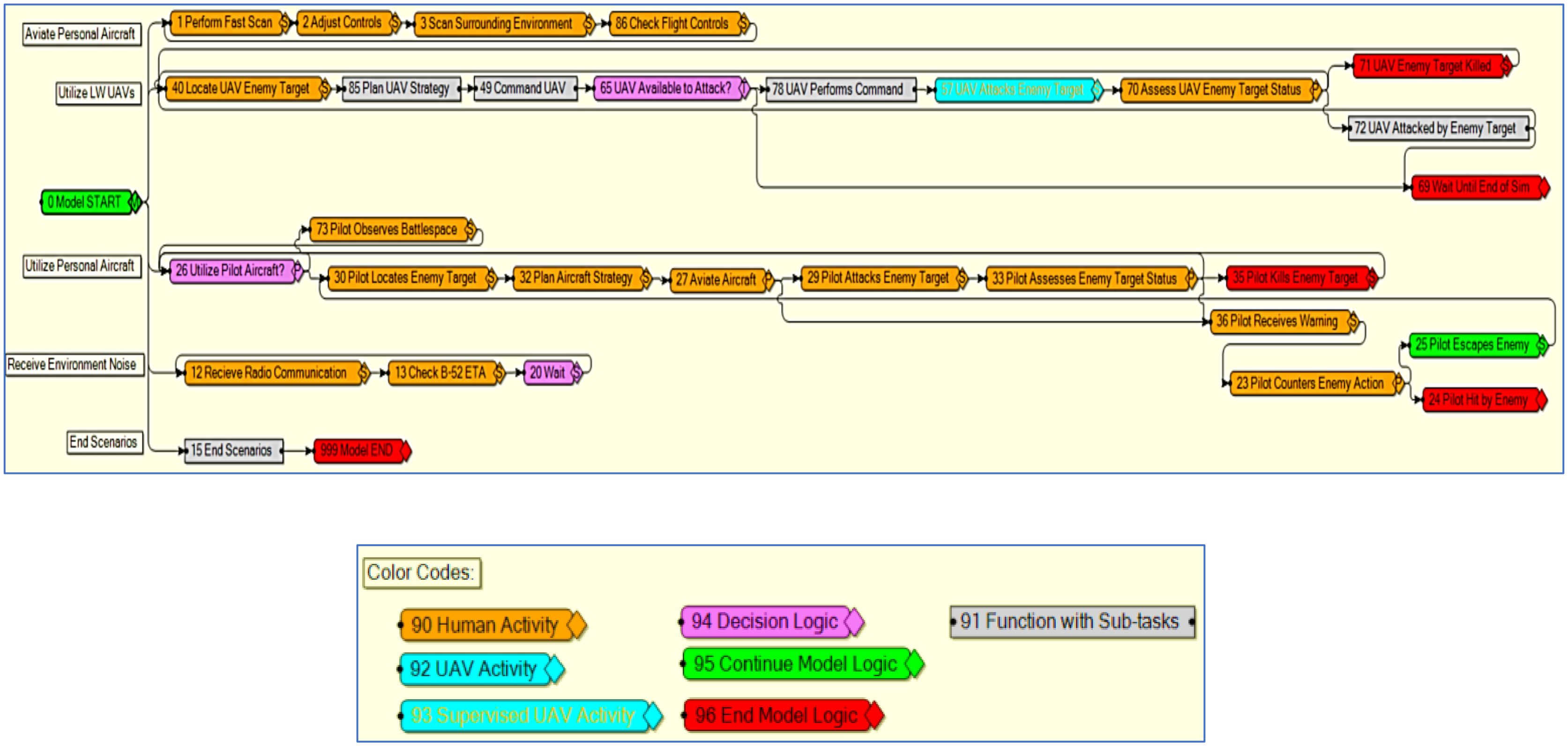


Figure 21. Overarching IMPRINT Task Network

Furthermore, some of the more complicated activities such as planning strategies, commanding the UAVs, and targeting adversary targets were decomposed into smaller sub-tasks. Figure 22-Figure 27 below illustrate some of the more complicated activities that were decomposed into smaller sub-tasks in IMPRINT.

In Figure 22, the "Plan UAV Strategy" function was broken down into workloads associated with controlling one or two UAVs to three or four UAVs.

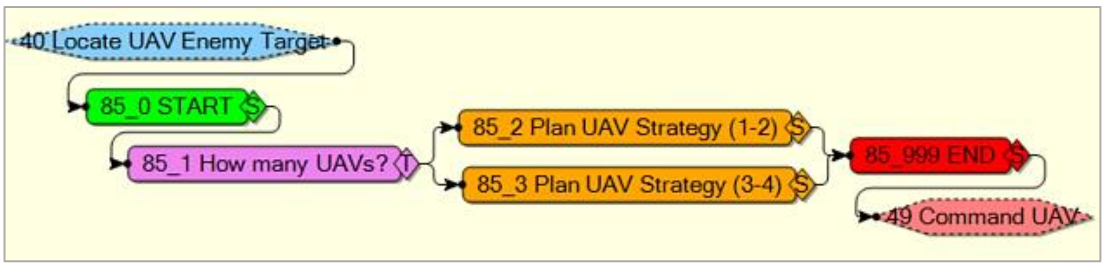


Figure 22. Plan UAV Strategy IMPRINT Function

In Figure 23, the "Command UAV" function was broken down into specific tasks the pilot would need to perform to command a single UAV.

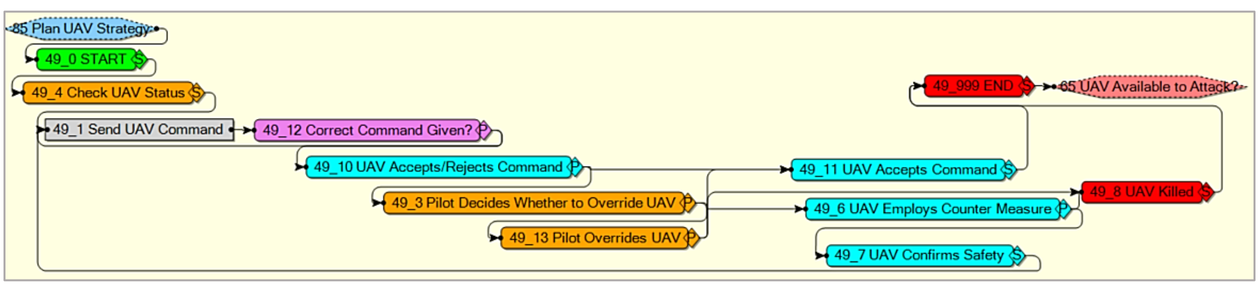


Figure 23. Command UAV IMPRINT Function

In Figure 24, the "Send UAV Command" sub-function was broken down into Tactical Battle Manager (high level of workload) commands, Pilot Directed Engagement (medium level of workload) commands, and Vector Steering (low level of workload) commands. Tactical Battle Manager commands utilize a higher level of automation to attack an adversary target. Pilot

Directed Engagement commands utilize a lower level of automation to execute formation or targeting actions. Finally, Vector Steering utilizes the lowest level of automation to follow pilot directed commands such as turning left or right as well as flying at a specific airspeed or heading.

Figure 24. Send UAV Command IMPRINT Sub-Function

In Figure 25, the “UAV Performs Command” function correlated to the level of command given by the pilot to a single UAV. The level of workload placed on the pilot increased as the level of autonomy decreased from Tactical Battle Manager to Vector Steering because lower level commands required a greater amount of manual control as well as mental processing for the Pilot to command a UAV. In addition, the amount of time it took the UAV to execute a pilot’s command decreased as the level of command decreased because a low level command was less complicated for the UAV to execute.

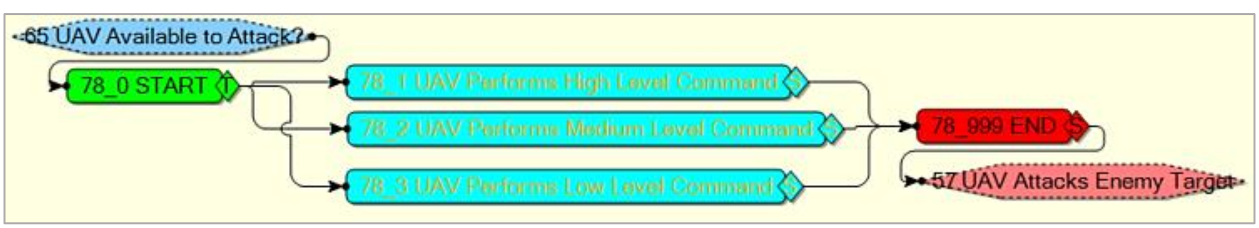


Figure 25. UAV Performs Command IMPRINT Function

In Figure 26, the “UAV Attacked by Enemy Target” function considered the case in which a UAV was attacked by an adversary and needed to evade from the enemy’s fire.

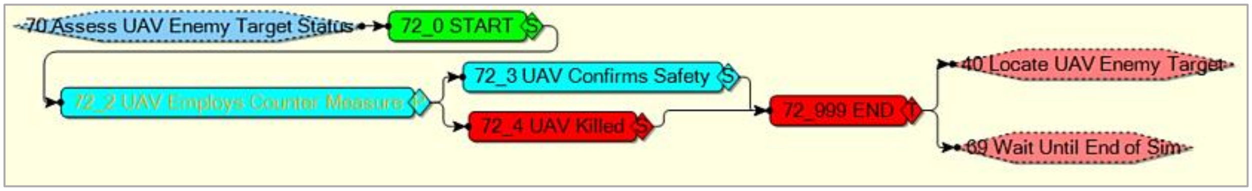


Figure 26. UAV Attacked by Enemy Target IMPRINT Function

In Figure 27, the “End Scenarios” function contained all four of the potential ending scenarios and the corresponding system logic for each case.

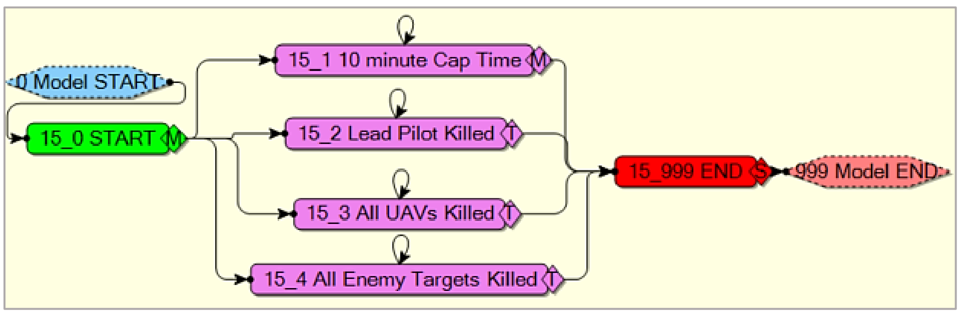


Figure 27. End Scenarios IMPRINT Function

Phase 3: Data Collection

The task network built in IMPRINT was then verified by SMEs who had experience developing and testing the virtual simulation cockpit in the ATACM study. The SMEs walked through the task network diagram for logical flow and gave predicted workload values based on the baseline model task descriptions and an explanation of VACP (Bierbaum et al., 1989). The individual tasks and their assigned values are listed in Table 17.

Table 17. IMPRINT Task Workload Demand Levels

	Task	Interface	Workload Demand					
			Auditory	*Cognitive*	*Fine Motor*	*Speech*	*Tactile*	*Visual*
Aviate Aircraft	Perform Fast Scan	Display		6.80				3.00
	Adjust Controls	Joystick		6.80	2.60		2.00	
	Scan Surrounding Environment	Display		6.80				4.40
	Check Flight Controls	Display		6.80				3.00
Utilize UAVs	Locate UAV Enemy Target	Display		6.80				4.40
	Plan UAV Strategy for 1 UAV	Display		6.80				4.00
	Plan UAV Strategy for 2 UAVs	Display		6.80				4.40
	Plan UAV Strategy for 3 UAVs	Display		6.80				5.00
	Check UAV Status	Display		4.60				4.40
	Initiate Call	Joystick		1.00	2.20		1.00	
	High Level Command (TBM)	Headset	4.30	4.60		2.00		
	Medium Level Command (PDE)	Headset	4.30	5.00		2.00		
	Low Level Command (VS)	Headset	4.30	5.30		2.00		
	Confirm Command	Joystick		1.00	2.20		1.00	
	Pilot Decides Whether to Override UAV	Display		6.80				4.40
	Pilot Overrides UAV	Headset	4.30	5.00		2.00		
	Pilot Overrides UAV	Joystick		1.00	2.20		1.00	
	UAV Performs High Level Command (TBM)	Display		6.80				4.00
	UAV Performs Medium Level Command (PDE)	Display		6.80				4.40
	UAV Performs Low Level Command (VS)	Display		6.80				5.00

	UAV Attacks Enemy Target	Display		4.60				3.00
	Assess UAV Enemy Target Status	Display		4.60				3.00
	Assess UAV Enemy Target Status	Headset	3.00	4.60		2.00		
	UAV Employs Counter Measure	Display		4.60				3.00
Utilize Personal Aircraft	Pilot Observes Battlespace	Display		6.80				4.40
	Pilot Locates Enemy Target	Display		4.60				4.00
	Plan Aircraft Strategy	Display		6.80				3.00
	Navigate Aircraft to Target Point	Display		6.80				4.40
	Navigate Aircraft to Target Point	Joystick		6.80	2.60		2.00	
	Pilot Attacks Enemy Target	Display		6.80				6.00
	Pilot Attacks Enemy Target	Joystick		6.80	4.60		2.00	
	Pilot Assesses Enemy Target Status	Display		4.60				3.00
	Pilot Assesses Enemy Target Status	Headset	3.00	4.60		2.00		
	Pilot Receives Warning	Display		1.00				3.00
	Pilot Receives Warning	Headset	3.00	1.00		2.00		
	Pilot Counters Enemy Action	Display		6.80				6.00
	Pilot Counters Enemy Action	Joystick		6.80	2.60		2.00	
Noise	Receive Radio Communication	Headset	3.00	4.60		2.00		
	Check B-52 ETA	Display	1.20					3.00

Once the task network was built and the workload values were inputted for each task, it was necessary to determine the probability and time that each task was expected to occur. The task probabilities and time distributions related to the successful completion or failure of certain tasks was calculated by extracting timing and decision data from the video footage of the nine test subjects in the ATACM study. The footage captured the pilots' audio commands, flight information shown on the Head-Down Display, and the time elapsed. The individual probabilities for specific task nodes are listed in Table 18-Table 25.

Table 18. Total number of Pilot Command Occurrences

Type	***Command Level***	***Total***
Attack	High	249
FormUp	Med	61
FormationNavigation	Med	10
TargetedNavigation	Med	29
WaypointNavigation	Low	43
FreeNavigation	Low	143

Table 19. Probability of Pilot Command Level

	Total Number	***Probability***
High	249	0.4654
Medium	100	0.1869
Low	186	0.3477

Table 20. Probability UAV Declined Command

	Total Number	***Probability***
Accepted	527	0.9777
Declined	12	0.0223

Table 21. Probability Pilot Overrode UAV

	Total Number	***Probability***
Overridden	15	0.6818
Not Overridden	7	0.3182

Table 22. Probability Pilot Repeated Command

	Total Number	***Probability***
Repeated	64	0.1192
Not Repeated	473	0.8808

Table 23. Survival Probabilities from UAV-Enemy Interactions

	UAV	***Enemy***
Killed	0.0432	0.5463
Survived	0.9568	0.4537

Table 24. Survival Probabilities from Pilot-Enemy Interactions

	Pilot	***Enemy***
Killed	0.2449	0.1122
Survived	0.7551	0.8878

Table 25. Probability Enemy Target Survived and Re-Attacked

	UAV	***Pilot***
Re-Attacked	0.0370	0.1312
No Re-Attack	0.4167	0.7565

Phase 4: Input Analysis

Upon completion of the data collection effort, input data modeling was performed on several aircraft aviation and targeting tasks in order to form probability distributions using ExpertFit software (Law, 2006). These probability distributions were tested for independence, homogeneity, and goodness-of-fit (see Figure 28Figure 33 on pages 108-112 and Table 26 on page 113). All of the final distributions in the baseline model either successfully passed these tests or were replaced by an empirical distribution directly representing the data. The analyzed input data was then synthesized with the task network diagram in IMPRINT to create the final baseline simulation model that featured the task flows, workload levels, system resources, probabilistic events, and process probability distributions.

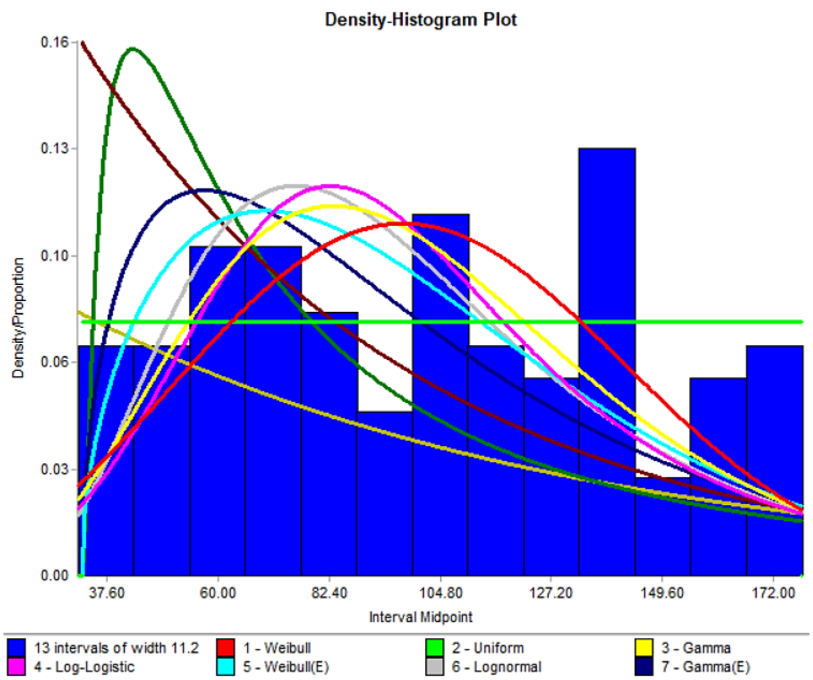


Figure 28. Probability Distribution Analysis of "UAV Performs High Level Command"

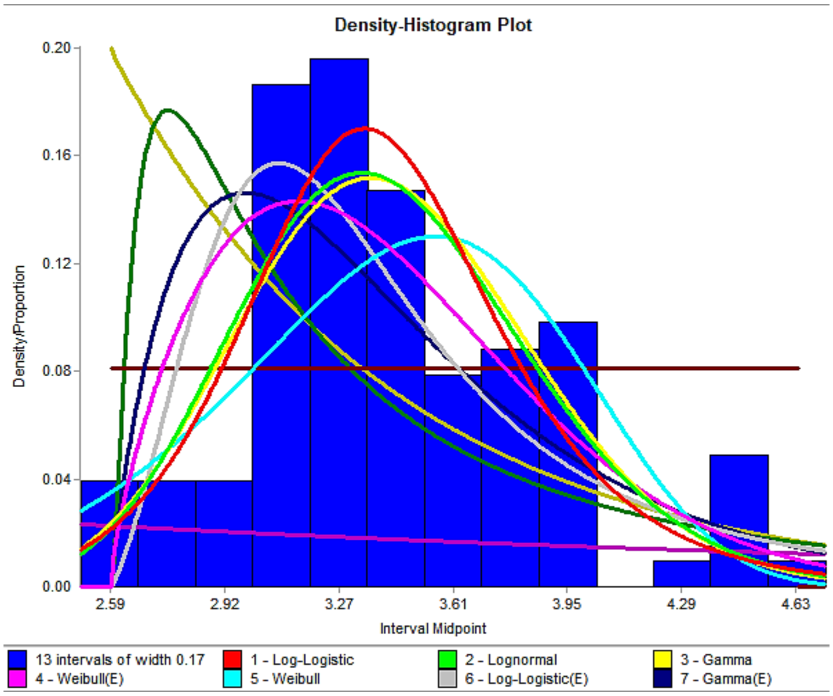


Figure 29. Probability Distribution Analysis of "UAV Performs Medium Level Command"

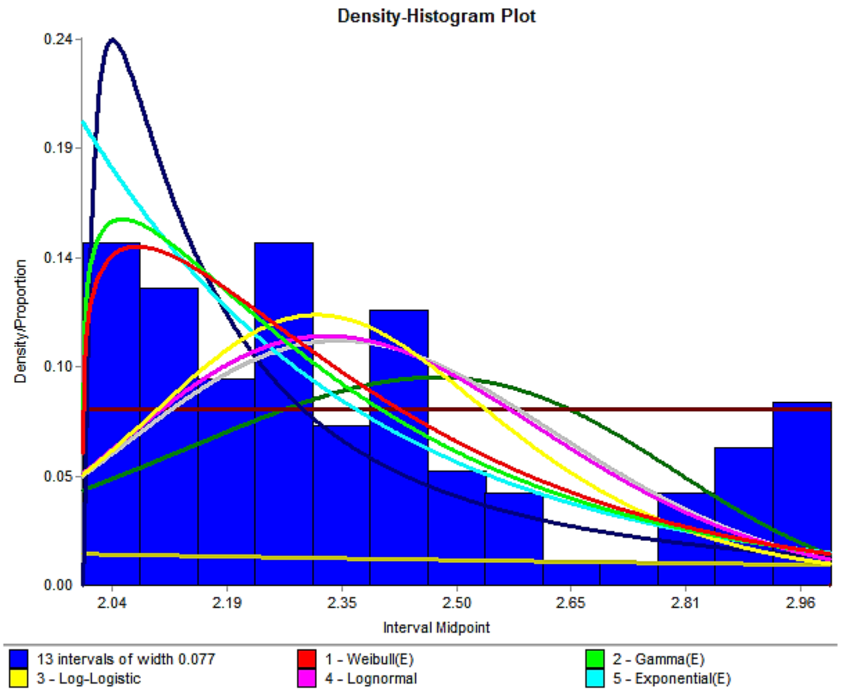


Figure 30. Probability Distribution Analysis of "UAV Performs Low Level Command"

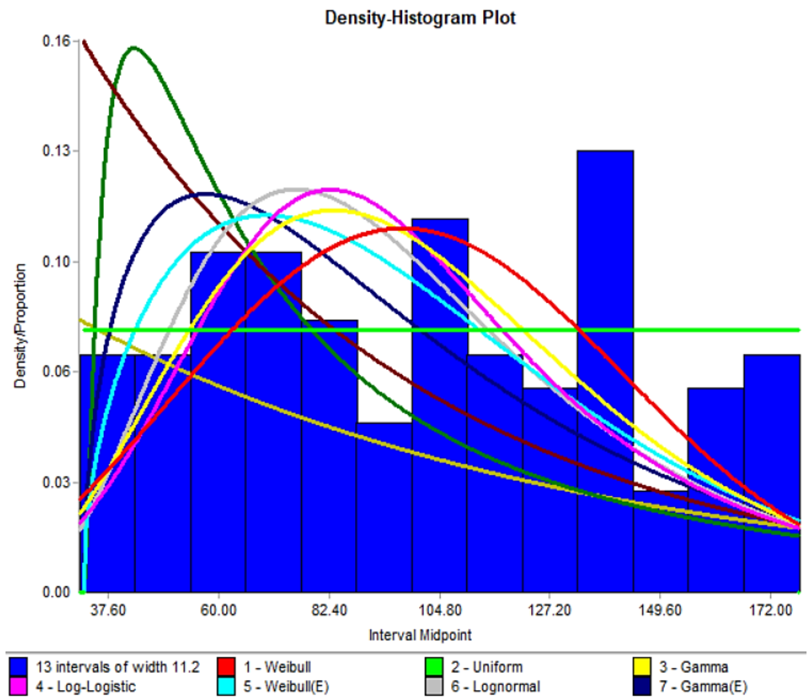


Figure 31. Probability Distribution Analysis of "UAV Attacks Enemy Target Analysis"

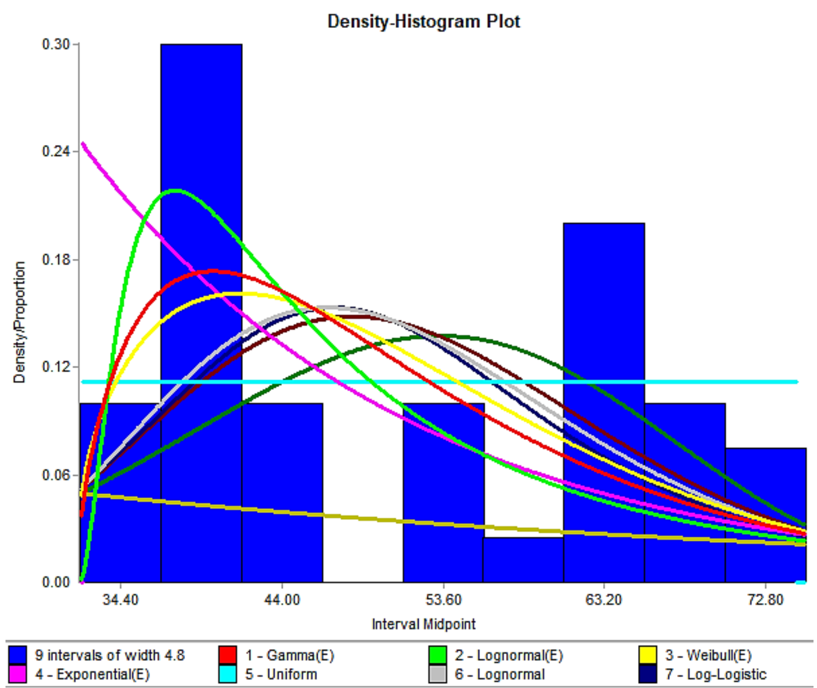


Figure 32. Probability Distribution Analysis of "UAV Employs Counter Measure"

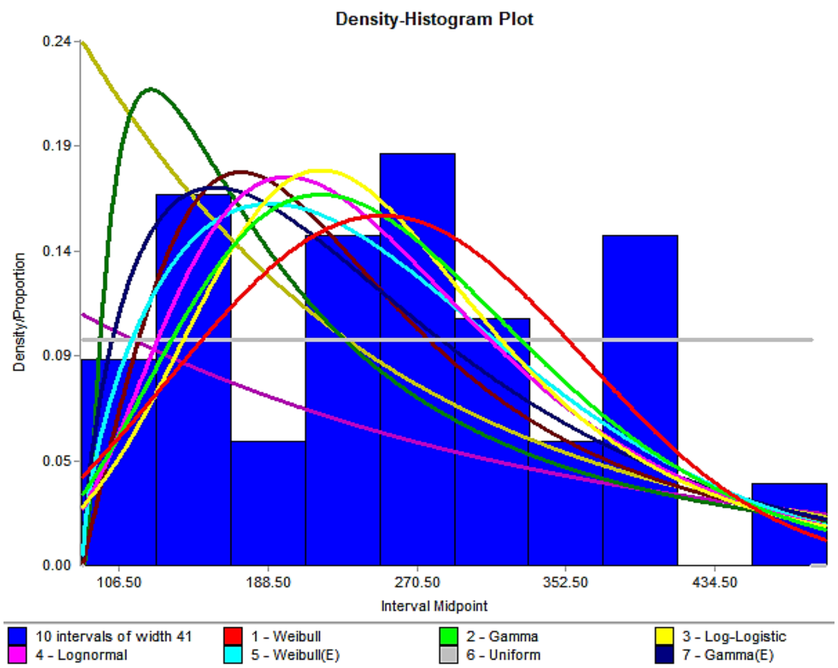


Figure 33. Probability Distribution Analysis of "Aviate Aircraft"

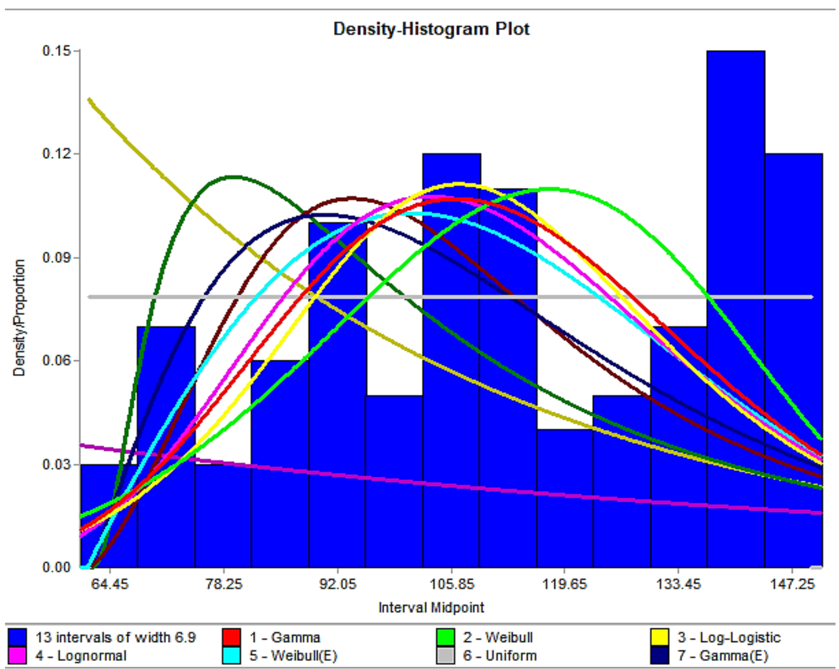


Figure 34. Probability Distribution Analysis of "Pilot Attacks Enemy Target"

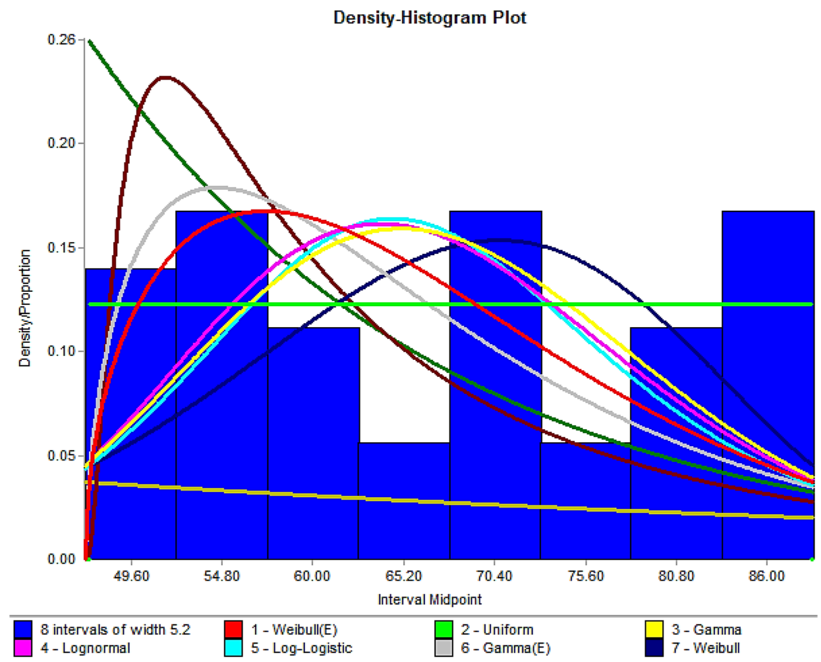


Figure 35. Probability Distribution Analysis of "Pilot Counters Enemy Action"

Table 26. Chi-Square Tests of Expert Fit Probability Distributions for Tasks 1-8

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8
Model	Weibull	Log-Logistic	Weibull	Weibull	Gamma	Weibull	Weibull	Gamma
Relative Score	91.67	95.00	97.22	91.67	83.33	97.50	83.33	92.50
Parameters	Location: 0.00 Scale: 114.64 Shape: 2.77	Location: 0.00 Scale: 3.38 Shape: 13.74	Location: 1.99 Scale: 0.40 Shape: 1.16	Location: 0.00 Scale: 114.64 Shape: 2.77	Location: 0.00 Mean: 51.63 STD: 13.50	Location: 0.00 Scale: 292.83 Shape: 2.85	Location: 47.04 Scale: 22.46 Shape: 1.47	Location: 0.00 Mean: 112.67 STD: 25.48
Mean Model Error	0.32%	0.31%	0.08%	0.32%	N/A	0.24%	0.31%	N/A
Model Evaluation	Borderline	Good	Good	Borderline	Borderline	Good	Good	Borderline

Phase 5: Validation of IMPRINT Model

Validation of the workload model was a key step in creating the baseline simulation model. This execution of this step provided the statistical evidence that the model sufficiently mirrored the real world system, which in this case was the ATACM study. To validate the DES, performance data and VACP values for workload were gathered as outputs from IMPRINT and compared to the results obtained from the ATACM study. Due to the low probability of achieving specific conditions such as the pilot repeating a command or the UAV declining a command, a total of 1,000 trials were run to ensure that each condition within the various task logic loops was achieved during the DES.

After running 1,000 trials in IMPRINT, the mission performance results were calculated by computing the percentage of total UAVs and the percentage of total enemy targets left at the end of each trial, as shown in Figure 36 and Figure 37. For satisfactory validation, an average absolute error that was within 10% was desired. According to the data, the mission performance varied between 1.04% average absolute error for the UAV survival results and 5.71% average absolute error for the enemy target survival results when comparing the IMPRINT model to the ATACM study.

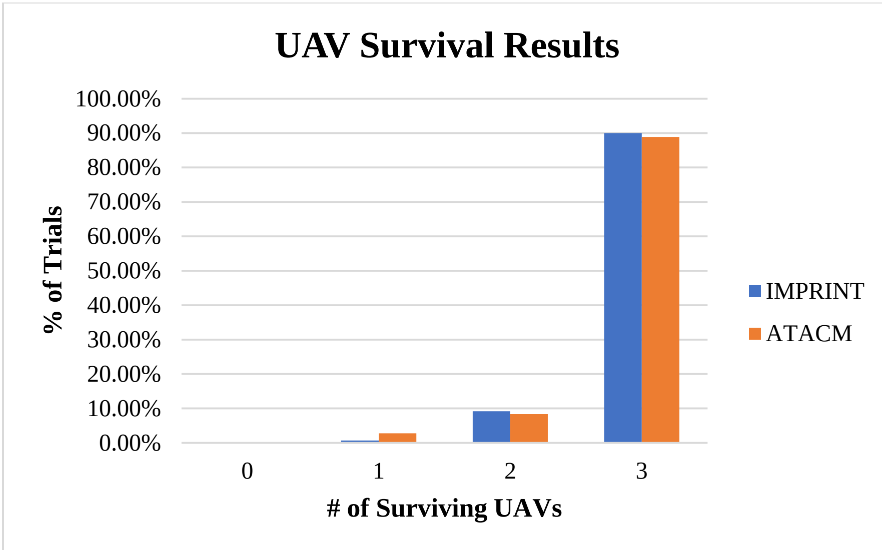


Figure 36. Validation Graph of UAV Survival Results

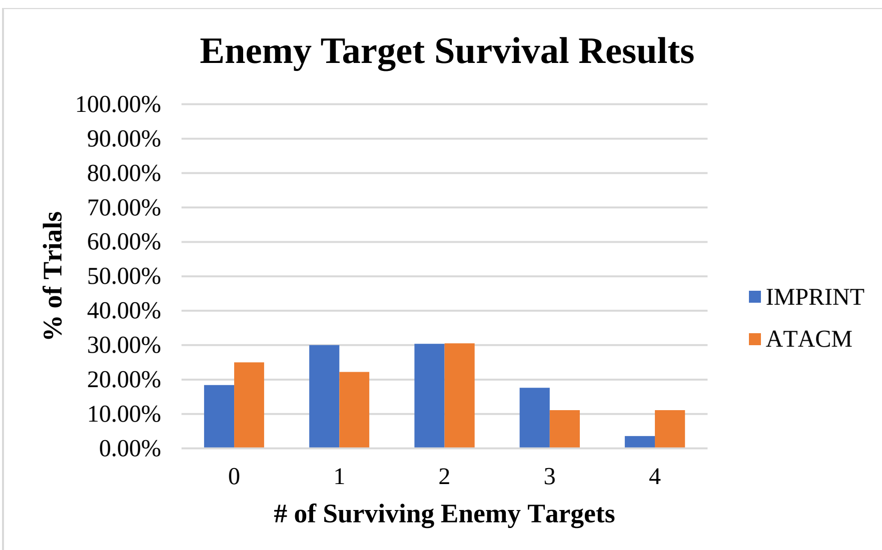


Figure 37. Validation Graph of Enemy Target Survival Results

The amount of time it took each simulation to run in IMPRINT was also compared to the length of time needed to complete each trial in the ATACM study. According to the graphs, shown in Figure 38 and Figure 39, the trials generally took about 9-10 minutes to complete for both studies. However, it should be noted that the IMPRINT model ran 1,000 simulations, while the ATACM study only performed 36 trials due to resource constraints. Despite the difference in total trials performed, the general trend of the IMPRINT performance times adequately reflected the overall tendency of the ATACM study.

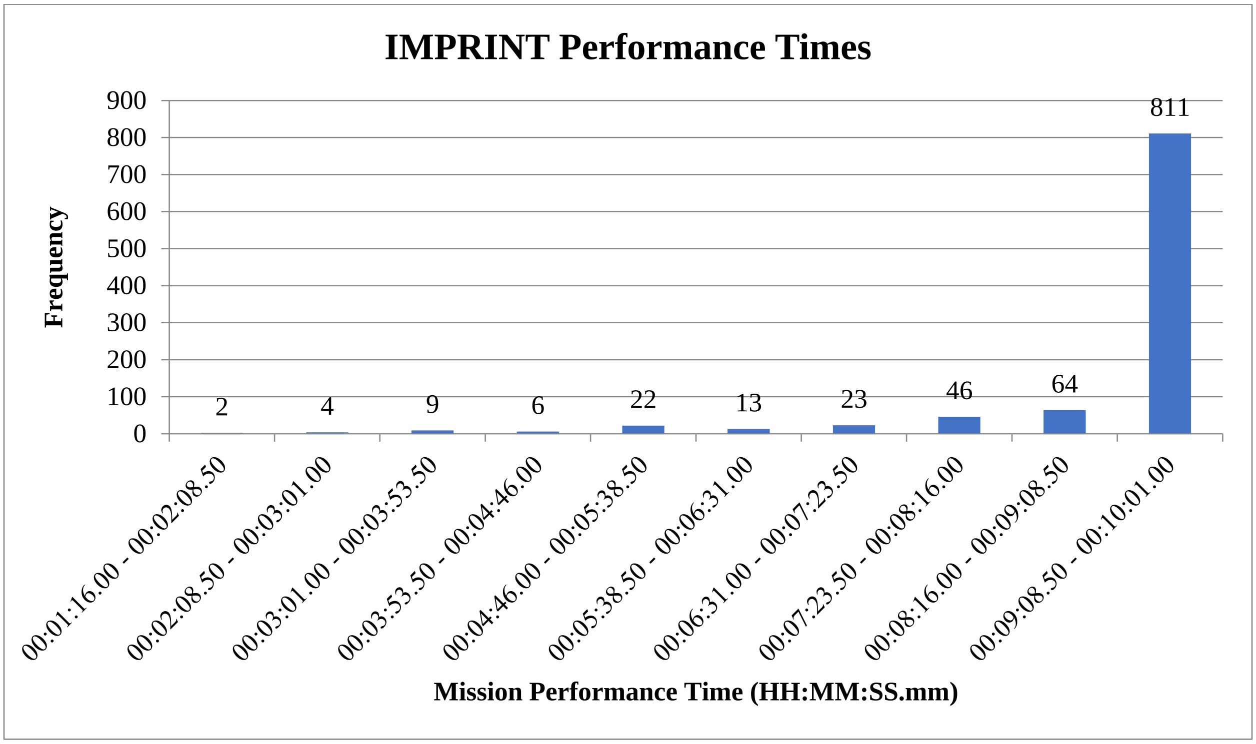


Figure 38. Histogram of IMPRINT Performance Times

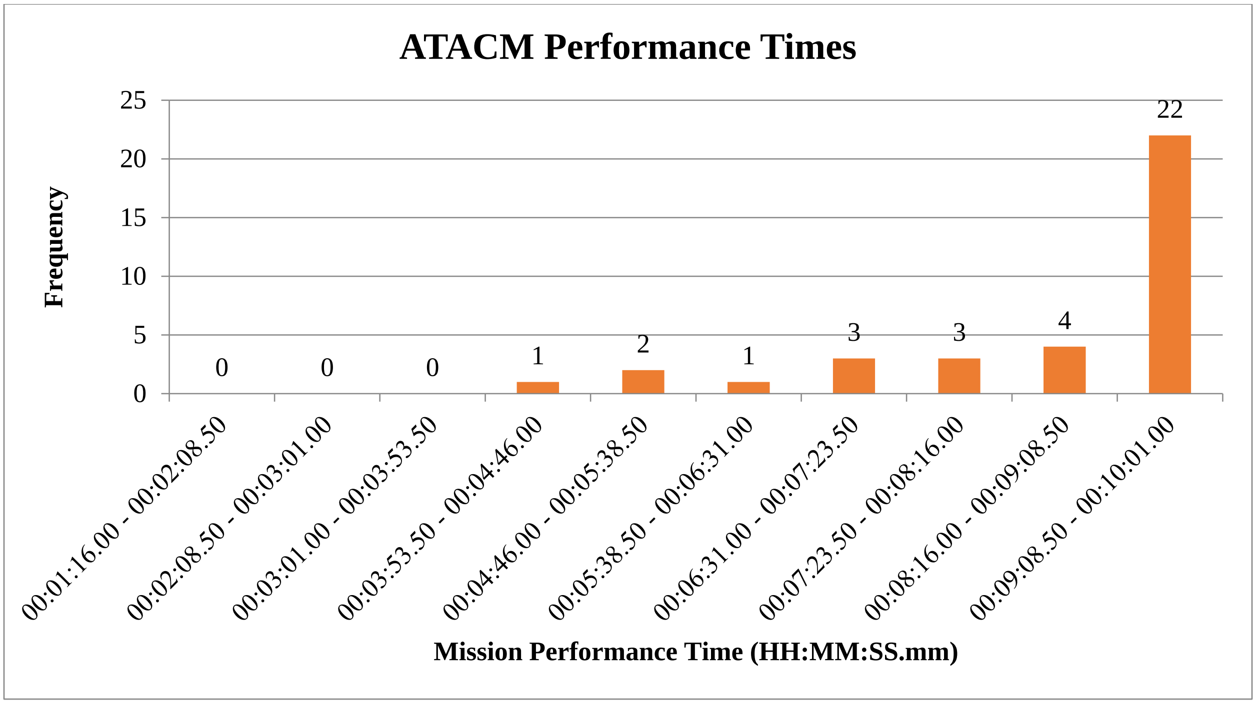


Figure 39. Histogram of ATACM Performance Times

For satisfactory validation, a confidence interval range that was within 10% above and below the mean was desired. For the ATACM trials, the average time in a given scenario was 8.58 minutes, thus a half-width of 0.86 min or less was required. A 99% confidence interval for this system produced a half-width of 0.85 minutes, thus a 99% confidence interval level was deemed sufficient for use in validation. The average time in the simulation was 9.42 minutes, which indicated that the simulation was on average 50.50 seconds slower than the study. It was hypothesized that the inability for the model to account for multiple attacks occurring in a short period of time is what instigated a slightly slower time in the system. Nonetheless, the overlap of both confidence intervals revealed that there was no statistical difference between the DES and the ATACM system, thus validating the IMPRINT model.

Appendix 4: ANOVA Tests and Tukey Groupings

Table 27 shows the results obtained from the one-way ANOVA test for the UAV survival rate data using a 95% confidence interval. According to the results, there is strong evidence against the null hypothesis since the p-value is less than 0.05. Therefore we reject the null hypothesis, which means that there is a definite, consequential relationship between the amount of autonomous control abstraction and the UAV survival rate.

Table 27. One-Way ANOVA Test for UAV Survival Rate Data using 95% Confidence Interval

Source	DF	Sum of Square	Mean Square	F Statistic	P-value
Groups *(between groups)*	3	2.456750	0.818917	7.776576	0.0000355626
Error *(within groups)*	3996	420.801010	0.105306		
Total	3999	423.257760	0.105841		

Table 28 shows the results obtained from the Tukey HSD test using 95% confidence interval for the UAV survival rate data. According to the results, VS-TBM, VS-Combination, and PDE-TBM pairs were statistically different from each other. However, there was not a statically significant difference among sample means for VS-PDE, PDE-Combination, and TBM-Combination pairs.

Table 28. Tukey HSD Test for UAV Survival Rate Data using 95% Confidence Interval

Pair	**Difference**	**SE**	**Q**	**Lower CI**	**Upper CI**	**Critical Mean**	**P-value**
VS-PDE	0.02200	0.01026	2.14386	-0.01530	0.05930	0.03730	0.42794
VS-TBM	0.06500	0.01026	6.33414	0.02770	0.10230	0.03730	0.00005
VS-Combo	0.04800	0.01026	4.67752	0.01070	0.08530	0.03730	0.00525
PDE-TBM	0.04300	0.01026	4.19028	0.00570	0.08030	0.03730	0.01620
PDE-Combo	0.02600	0.01026	2.53366	-0.01130	0.06330	0.03730	0.27748
TBM-Combo	0.01700	0.01026	1.65662	-0.02030	0.05430	0.03730	0.64499

Table 29 shows the results obtained from the one-way ANOVA test for the UAV survival rate data using a 95% confidence interval. According to the results, there is strong evidence against the null hypothesis since the p-value is less than 0.05. Therefore we reject the null hypothesis, which means that there is a definite, consequential relationship between the amount of autonomous control abstraction and the enemy target survival rate.

Table 29. One-Way ANOVA Test for Enemy Target Survival Rate Data using 95% Confidence Interval

Source	**DF**	**Sum of Square**	**Mean Square**	**F Statistic**	**P-value**
Groups *(between groups)*	4	3388.780800	847.195200	1088.124913	0.00000
Error *(within groups)*	4995	3889.020435	0.778583		
Total	4999	7277.801235	1.455851		

Table 30**Error! Reference source not found.** shows the results obtained from the Tukey HSD test using 95% confidence interval for the enemy target survival rate data. According to the results, all of the pairs were statistically different from each other.

Table 30. Tukey HSD Test for UAV Survival Rate Data using 95% Confidence Interval

Pair	Difference	SE	Q	Lower CI	Upper CI	Critical Mean	p-value
Fully Manual-VS	1.312000	0.027903	47.019875	1.204320	1.41968	0.10768	0.000000
Fully Manual-PDE	2.004000	0.027903	71.819992	1.896320	2.11168	0.10768	0.000000
Fully Manual-TBM	2.359000	0.027903	84.542595	2.251320	2.46668	0.10768	0.000000
Fully Manual-Combo	1.834000	0.027903	65.727477	1.726320	1.94168	0.10768	0.000000
VS-PDE	0.692000	0.027903	24.800117	0.584320	0.79968	0.10768	0.000000
VS-TBM	1.047000	0.027903	37.522720	0.939320	1.15468	0.10768	0.000000
VS-Combo	0.522000	0.027903	18.707603	0.414320	0.62968	0.10768	0.000000
PDE-TBM	0.355000	0.027903	12.722603	0.247320	0.46268	0.10768	0.000000
PDE-Combo	0.170000	0.027903	6.092514	0.062320	0.27768	0.10768	0.000163
TBM-Combo	0.525000	0.027903	18.815118	0.417320	0.63268	0.10768	0.000000

Bibliography

88 Air Base Wing Public Affairs. (2019). XQ-58A Valkyrie Demonstrator Completes Inaugural Flight. Retrieved November 12, 2019, from U.S. Air Force website: https://www.wpafb.af.mil/News/Article-Display/Article/1777743/xq-58a-valkyrie-demonstrator-completes-inaugural-flight/

Adams, M. J., & Pew, R. W. (1990). Situational Awareness in the Commercial Aircraft Cockpit: A Cognitive Perspective. *9th IEEE/AIAA/NASA Conference on Digital Avionics Systems*, 519–524. Retrieved from https://ieeexplore-ieee-org.afit.idm.oclc.org/document/111342%0D

Alion Science and Technology Corporation. (2009). *Improved Performance Research Integration Tool (IMPRINT)*. Retrieved from https://www.arl.army.mil/www/pages/446/IMPRINTPro_vol2.pdf

Allen, J. E., Guinn, C. I., & Horvitz, E. (1999). Mixed-Initiative Interaction. *IEEE Intelligent Systems*, *14*(5), 14–23.

Allender, L. (2000). Modeling Human Performance: Impacting System Design, Performance, and Cost. *Military, Government and Aerospace Simulation Symposium, 2000 Advanced Simulation Technologies Conference*, *32*(3), 139–144. Retrieved from http://www.arl.army.mil/www/pages/447/Astc2000-Allender.pdf

Andrews, J. M. (2020). *A New Model of Airpower: Development of an IMPRINT Model to Analyze the Effects of Manned Unmanned Teaming on Mental Workload*. Air Force Institute of Technology, Wright Patterson Air Force Base.

Antsaklis, P. J., Passino, K. M., & Wang, S. J. (1991, June). An Introduction to Autonomous Control Systems. *IEEE Control Systems Magazine*, 5–12.

Baddeley, A. D. (1996). *Working Memory*. Oxford, UK: Clarendon.

Beevis, D. (1992). *Analysis Techniques for Man-Machine Systems Design: A Report Produced Under the Auspices of NATO Defense Research Group Panel 8*. Ontario, Canada.

Bierbaum, C. R., Szabo, S. M., & Aldrich, T. B. (1989). Task Analysis of the UH-60 Mission and Decision Rules for Developing a UH-60 Workload Prediction Model. *US Army Research Institute*.

Billings, C. E. (1991). Human-Centered Aircraft Automation: A Concept and Guidelines. In *NASA*. Retrieved from https://ntrs.nasa.gov/search.jsp?R=19910022821

Broadbent, D. E. (1958). *Perception and Communication* (1st ed.). Retrieved from https://www.elsevier.com/books/perception-and-communication/broadbent/978-1-4832-0079-8

Cain, B. (2004). A Review of the Mental Workload Literature. *Defence Research and Development Canada*, (1998), 1–34. Retrieved from https://www.researchgate.net/publication/235159082_A_Review_of_the_Mental_Workload_Literature

Cassenti, D. N., & Kelley, T. D. (2006). Towards the shape of mental workload. *Proceedings of the Human Factors and Ergonomics Society*, 1147–1151. https://doi.org/10.1177/154193120605001107

Cassenti, D. N., Kelley, T. D., Colle, H. A., & McGregor, E. A. (2011). Modeling Performance Measures and Self-Ratings of Workload in a Visual Scanning Task. *Human Factors and Ergonomics Society 55th Annual Meeting*, 870–874. https://doi.org/10.1177/1071181311551181

Childress, M., Hart, S., & Bortolussi, M. (1982). The Reliability and Validity of Flight Task

Workload Ratings. *Human Factors Society 26th Annual Meeting*, 319–323. https://doi.org/10.1177/154193128202600411

Clare, A. S., Maere, P. C. P., & Cummings, M. L. (2012). Assessing Operator Strategies for Real-Time Replanningo of Multiple Unmanned Vehicles. *Intelligent Decision Technologies*, *6*(3), 221–231. https://doi.org/10.3233/IDT-2012-0138

Clough, B. T. (2002). Metrics, Schmetrics! How the Heck Do You Determine a UAV's Autonomy Anyway? *Performance Metrics for Intelligent Systems Workshop*. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA515926

Colombi, J. M., Miller, M. E., Schneider, M., Mcgrogan, M. J., Long, C. D. S., & Plaga, J. (2011). Predictive Mental Workload Modeling for Semiautonomous System Design: Implications for Systems of Systems. https://doi.org/10.1002/sys

Cummings, M. L., Bruni, S., Mercier, S., & Mitchell, P. J. (2007). Automation Architecture for Single Operator, Mulitple UAV Command and Control. *The International C2 Journal*, *1*(2), 1–24.

Curry, R. E., Jex, H., Levison, W., & Stassen, H. (1979). Final Report of the Control Engineering Group. In *Moray N. (eds) Mental Workload* (pp. 235–254). Retrieved from https://doi-org.afit.idm.oclc.org/10.1007/978-1-4757-0884-4_13

De Visser, E. J., Legoullon, M., Freedy, A., Freedy, E., Weltman, G., & Parasuraman, R. (2008). Designing an Adaptive Automation System for Human Supervision of Unmanned Vehicles: A Bridge from Theory to Practice. *Proceedings of the Human Factors and Ergonomics Society*, *1*, 221–225. https://doi.org/10.1177/154193120805200405

De Waard, D. (1996). The Measurement of Drivers ' Mental Workload (University of

Groningen, Haren, The Netherlands). Retrieved from http://dissertations.ub.rug.nl/FILES/faculties/ppsw/1996/d.de.waard/09_thesis.pdf

Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., & Zoladz, P. R. (2007). The Temporal Dynamics Model of Emotional Memory Processing: A Synthesis on the Neurobiological Basis of Stress-Induced Amnesia, Flashbulb and Traumatic memories, and the Yerkes-Dodson law. *Neural Plasticity*, *2007*. https://doi.org/10.1155/2007/60803

Donmez B., Nehme C., & Cummings M.L. (2010). Modeling Workload Impact in Multiple Unmanned Vehicle Supervisory Control. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *40*(5), 1180–1190. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5484493

Draper, J. V. (1995). Teleoperators for advanced manufacturing: applications and human factors challenges. *The International Journal of Human Factors in Manufacturing*, *5*(1), 53–85. https://doi.org/10.1002/hfm.4530050105

Drew, J. (2016). Pentagon Touts “Loyal Wingman” for Combat Jets. Retrieved November 13, 2018, from Flight Global website: https://www.flightglobal.com/news/articles/pentagon-touts-loyal-wingman-for-combat-jets-423682/

Endsley, M. R. (1987). The Application of Human Factors to the Development of Expert Systems for Advanced Cockpits. *Human Factors Society 31st Annual Meeting*, *31*(12), 1388–1392. Retrieved from https://journals-sagepub-com.afit.idm.oclc.org/doi/pdf/10.1177/154193128703101219

Endsley, M. R. (1996). Automation and Situation Awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 163–181). Retrieved from https://www.crcpress.com/Automation-and-Human-Performance-Theory-and-Applications/Parasuraman-Mouloua/p/book/9780805816167

Endsley, M. R., & Kaber, D. B. (1999a). Level of Automation Effects on Performance, Situation Awareness, and Workload in a Dynamic Control Task. *Ergonomics*, *42*(3), 462–492.

Endsley, M. R., & Kaber, D. B. (1999b). Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task. *Ergonomics*, *42*(3), 462–492. https://doi.org/10.1080/001401399185595

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Fereidunian, A., Lehtonen, M., Lesani, H., Lucas, C., & Nordman, M. (2007). Adaptive Autonomy: Smart Cooperative Cybernetic Systems for More Humane Automation Solutions. *IEEE International Conference on Systems, Man and Cybernetics*, 202–207. https://doi.org/10.1109/ICSMC.2007.4413874

Fereidunian, A., Lucas, C., Lesani, H., Lehtonen, M., & Nordman, M. (2007). Challenges in Implementation of Human-Automation Interaction Models. *Mediterranean Conference on Control and Automation*, 1–6. https://doi.org/10.1109/MED.2007.4433895

Fong, T. W., Thorpe, C., & Baur, C. (1999). Collaborative Control: A Robotic-Centric Model for Vehicle Teleoperation. *AAAI 1999 Spring Symposium: Agents with Adjustable Autonomy*. Pittsburgh, PA: Carnegie Mellon University Robotics Intitute.

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. Retrieved from https://doi.org/10.1016/j.cogsys.2017.02.007

Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & St. John, M.

(2008). The Red-line of Workload: Theory, Research, and Design. *Human Factors and Ergonomics Society*, *2*, 1204–1208. https://doi.org/10.1177/154193120805201811

Groover, M. (2015). *Fundamentals of Modern Manufactoring: Materials, Processes, and Systems* (5th ed.). New York, NY: Wiley.

Hamilton, D. B., & Bierbaum, C. R. (1992). *Operator Workload Predictions for the Revised AH-64A Workload Prediction Model*. Retrieved from https://www.semanticscholar.org/paper/Operator-Workload-Predictions-for-the-Revised-I%3A-Hamilton-Bierbaum/8d8d46f4aea2ff4818fbbe35b2233764b6532fe8

Hancock, P. A., & Chignell, M. H. (1988). Mental Workload Dynamics in Adaptive Interface Design. *IEEE Transactions on Systems, Man and Cybernetics*, *18*(4), 647–658. https://doi.org/10.1109/21.17382

Hanlon, M. (2017). Kratos to Show Low-Cost Valkyrie and Mako “Wingman” Combat Drones. Retrieved November 13, 2018, from New Atlas website: https://newatlas.com/kratos-valkyrie-mako-combat-drones-paris/50044/

Harriott, C. E., Zhang, T., & Adams, J. A. (2013). Assessing Physical Workload for Human-Robot Peer-Based Teams. *International Journal of Human Computer Studies*, *71*(7–8), 821–837. https://doi.org/10.1016/j.ijhcs.2013.04.005

Harriott, C. E., Zhuang, R., Adams, J. A., & Deloach, S. A. (2012). Towards Using Human Performance Moderator Functions in Human-Robot Teams. *Human-Agent Interaction Design and Models*. Retrieved from https://www.semanticscholar.org/paper/Towards-Using-Human-Performance-Moderator-Functions-Harriott-Zhuang/f2a7e2afb64151ecb44c5ebdebc3f23dddb704b7

Hart, S. G., & Sheridan, T. B. (1984). Pilot Workload, Performance, and Aircraft Control

Automation. *AGARD Symposium on Human Factors Considerations in High Performance Aircraft*, 18.1-18.12. Retrieved from https://www.researchgate.net/publication/23888817_Pilot_workload_performance_and_aircraft_control_automation

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*, 139–183.

Hebb, D. O. (1955). Drives and the C.N.S. (Conceptual Nervous System). *Psychology Review*, *62*(4), 243–254.

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton: CRC Press.

Hugo, J., & Gertman, D. (2012). The Use of Computational Human Performance Modeling as Task Analysis Tool. *8th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies*, *1*, 90–101.

Hunn, B. P., & Heuckeroth, O. H. (2006). A Shadow Unmanned Aerial Vehicle (UAV) Improved Performance Research Integration Tool (IMPRINT) Model Supporting Future Combat Systems. In *Department of Defense*. Retrieved from https://www.semanticscholar.org/paper/A-Shadow-Unmanned-Aerial-Vehicle-(UAV)-Improved-Hunn-Heuckeroth/5b451d2661dcb6c0c9db0d12af77218982ebea29

Joe, R. (2019, November). China's Growing High-End Military Drone Force. *The Diplomat*. Retrieved from https://thediplomat.com/2019/11/chinas-growing-high-end-military-drone-force/

Johnson, C. D., Miller, M. E., Rusnock, C. F., & Jacques, D. R. (2017). A Framework for Understanding Automation in Terms of Levels of Human Control Abstraction. *IEEE*

International Conference on Systems, Man and Cybernetics, 1145–1150. Banff Canada.

Johnson, M., Bradshaw, J. M., & Feltovich, P. J. (2018). Tomorrow's Human–Machine Design Tools: From Levels of Automation to Interdependencies. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 77–82. https://doi.org/10.1177/1555343417736462

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014a). Coactive Design: Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction*, *3*(1), 43. https://doi.org/10.5898/jhri.3.1.johnson

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014b). Coactive Design: Designing Support for Interdependence in Joint Activity. *Journal of Human-Robot Interaction*, *3*(1), 43. https://doi.org/10.5898/jhri.3.1.johnson

Johnson, M., Bradshaw, J. M., Feltovich, P., Jonker, C., Van Riemsdijk, B., & Sierhuis, M. (2012). Autonomy and Interdependence in Human-Agent-Robot Teams. *IEEE Intelligent Systems*, *27*(2), 43–51. https://doi.org/10.1109/MIS.2012.1

Johnson, M., Vignati, M., & Duran, D. (2018). Understanding Human-Autonomy Teaming through Interdependence Analysis. *Symposium on Human Autonomy Teaming*, 1–20.

Johnson, R. D., Bershader, D., & Leifer, L. (1983). *Autonomy and the Human Element in Space: Executive Summary* (1st ed.). Retrieved from https://vufind.carli.illinois.edu/vf-gsu/Record/uiu_3966493

Kaber, D. B. (2018). Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation. *Journal of Cognitive Engineering and Decision Making*, *12*(1), 7–24. https://doi.org/10.1177/1555343417737203

Kaber, D. B., & Endsley, M. R. (2004). The Effects of Level of Automation and Adaptive Automation on Human Performance, Situation Awarenss, and Workload in a Dynamic

Control Task. *Theoretical Issues in Ergonomics Science*, *5*(2), 113–153. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/1463922021000054335

Kaber, D. B., Stoll, N., & Thurow, K. (2007). Human-Automation Interaction Strategies for Life Science Applications: Implications and Future Research. *Proceedings of the 3rd IEEE International Conference on Automation Science and Engineering*, 615–620. https://doi.org/10.1109/COASE.2007.4341856

Kahneman, D. (1973). *Attention and Effort*. https://doi.org/10.2307/1421603

Keller, J. (2002). Human Performance Modelling for Discrete-Event Simulation: Workload. *2002 Winter Simulation Conference*, *1*, 157–162. https://doi.org/10.1109/WSC.2002.1172879

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten Challenges for Making Automation a “Team Player” in Joint Human–Agent Activity. *IEEE Intelligent Systems*, *19*, 91–95. Retrieved from http://dx.doi.org/10.1109/MIS.2004.74

Law, A. M. (2006). *ExpertFit Analysis User Guide*. Tuscan, AZ: Averill M. Law & Associates.

Lorenz, B., Nocera, F. D., Röttger, S., Di Nocera, F., Rottger, S., & Parasuraman, R. (2001). The Effects of Level of Automation on the Out-of-the-Loop Unfamiliarity in a Complex Dynamic Fault-Management Task during Simulated Spaceflight Operations. *Human Factors and Ergonomics Society*, *45*(2), 44–48. https://doi.org/10.1177/154193120104500209

Magnuson, S. (2018). The Future of Air Power: New Age of Autonomous Jet Fighters On Horizon. *National Defense*, *103*(778), 30–32. Retrieved from http://eds.a.ebscohost.com.afit.idm.oclc.org/eds/detail/detail?vid=0&sid=ddf5f8a4-8165-42a7-8e23-

29318a5007e1%40sessionmgr4008&bdata=JnNpdGU9ZWRzLWxpdmU%3D#AN=131710160&db=a9h

McCracken, J. H., & Aldrich, T. B. (1984). Analyses of Selected LHX Mission Functions: Implications for Operator Workload and System Automation Goals. In *Aviation Research and Development Activity*. https://doi.org/10.21236/ADA232330

Milgram, P., Rastogi, A., & Grodski, J. J. (1995). Telerobotic Control Using Augmented Reality. *4th IEEE International Workshop on Robot and Human Communication*, 21–29. https://doi.org/10.1109/ROMAN.1995.531930

Miller, C. A. (2017). The FireFox Fallacy: Why Intent Should Be an Explicit Part of the External World in Human Automation Interaction. In P. J. Smith & R. R. Hoffman (Eds.), *Cognitive Systems Engineering: The Future for a Changing World* (1st ed., pp. 269–294). https://doi.org/10.1201/9781315572529

Mitchell, D. K., Samms, C., & Wojcik, T. M. (2006). System-of-systems Modeling: The Evolution of an Approach for True Human System Integration. *15th Conference on Behavior Representation in Modeling and Simulation*, 67–74.

Mitchell, D. K. (2009). Workload Analysis of the Crew of the Abrams V2 SEP : Phase I Baseline IMPRINT Model. In *Department of Defense*. Retrieved from https://www.arl.army.mil/arlreports/2009/technical-report.cfm?id=1883

Mitchell, D.K. (2000). Mental Workload and ARL Workload Modeling Tools. *U.S. Army Research Laboratory*.

Mitchell, Diane. K. (2003). *Advanced Improved Performance Research Integration Tool (IMPRINT) Vetronics Technology Test Bed Model Development*. Aberdeen Proving Ground, MD.

NASA. (1986). NASA Task Load Index (TLX). In *Human Performance Research Group*. https://doi.org/10.1007/springerreference_183995

National Research Council. (1982). *Automation in Combat Aircraft*. Retrieved from https://www.nap.edu/catalog/19605/automation-in-combat-aircraft%0D

Neerincx, M. A. (2003). Cognitive Task Load Analysis: allocation Tasks and Designing Support. *Handbook of Cognitive Task Design*, *0*, 283–305. https://doi.org/10.1007/s13398-014-0173-7.2

Norman, D. A. (1989). *The Problem of Automation: Inappropriate Feedback and Interaction, Not Overautomation*. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19900004678.pdf

Ntuen, C. A., & Park, E. H. (1988). Human Factors Issues in Teleoperated Systems. *First International Conference on Ergonomics of Hybrid Automated Systems I*, 203–311. Retrieved from http://dl.acm.org/citation.cfm?id=58235

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A Model For Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans: A Publication of the IEEE Systems, Man, and Cybernetics Society*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Pomranky, R. a, & Wojciechowski, J. Q. (2007). *Determination of Mental Workload During Operation of Multiple Unmanned Systems*.

Pop, V. L., Michelson, W. S., & Engineering, H. S. (2018). Human Workload Modeling for

Autonomous Ground Vehicles. *Ground Vehicle Systems Engineering and Technology Symposium (GVSETS)*. Retrieved from http://gvsets.ndia-mich.org/publication.php?documentID=643

Powers, D., & Gacy, M. (2018). Using IMPRINT to Model Operator Staffing an Workload Consideration in a 24/7 Full-Service Mission Operations Center. *15th International Conference on Space Operations*, (June), 1–12. https://doi.org/10.2514/6.2018-2394

Proud, R. W., Hart, J. J., & Mrozinski, R. B. (2003). Methods for Determining the Level of Autonomy to Design into a Human Spaceflight Vehicle: A Function Specific Approach. *Performance Metrics for Intelligent Systems*. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA515467

Reid, G. B., & Colle, H. A. (1988). Critical SWAT Values for Predicting Operator Overload. *Proceedings of the Human Factors Society Annual Meeting*, *32*(19), 1414–1418. https://doi.org/10.1177/154193128803201923

Riley, V. (1989). A General Model of Mixed-Initiative Human-Machine Systems. *Human Factors and Ergonomics Society Annual Meeting*, *33*(2), 124–128. https://doi.org/10.1177/154193128903300227

Rogoway, T. (2017). More Details Emerge On Kratos' Optionally Expendable Air Combat Drones. Retrieved November 13, 2018, from The War Zone Wire website: http://www.thedrive.com/the-war-zone/7449/more-details-on-kratos-optionally-expendable-air-combat-drones-emerge

Rosenberg, N. (1982). *Inside the Black Box: Technology and Economics* (1st ed.). Cambridge UK: Cambridge University Press.

Rusnock, C. F., & Geiger, C. D. (2013). Using Discrete-Event Simulation for Cognitive

Workload Modeling and System Evaluation. *2013 Industrial and Systems Engineering Research Conference A. Krishnamurthy*, (January). Orlando, FL: Research Gate.

Rusnock, C. F., & Geiger, C. D. (2014). Simulation-Based Assessment of Performance-Workload Tradeoffs for System Design Evaluation. *Proceedings of the 2014 Industrial and Systems Engineering Research Conference*, (May).

Schneider, M., & McGrogan, J. (2011). *Architecture Based Workload Analysis of UAS Multi-Aircraft Control: Implications of Implementation on MQ-1B Predator*.

Schumacher, C., Patzek, M., Aha, D., Falco, R., Novak, M., Orr, H., … Van Pelt, J. (2017). *Autonomy for Air Combat Missions (ATACM): Autonomy Research Pilot Initiative (ARPI) Report*. Wright Patterson Air Force Base.

Secarea, V. V. J. (1990). Beyond Knobs and Dials: Toward an Intentional Model of Man-Machine Interaction. *IEEE Conference on Aerospace and Electronics*, *2*, 763–769. Retrieved from https://ieeexplore-ieee-org.afit.idm.oclc.org/document/112864

Secretary of the Air Force Public Affairs. (2016). Flight Plan Outlines Next 20 Years for RPA. Retrieved November 12, 2019, from U.S. Air Force website: https://www.af.mil/News/Article-Display/Article/774728/flight-plan-outlines-next-20-years-for-rpa/

Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators. In *Office of Naval Research*. Retrieved from https://ntrs.nasa.gov/search.jsp?R=19790007441

Stevenson, B. (2019, June). "Loyal Wingman" Part of the Future of Air Combat. *Aviation International News*. Retrieved from https://www.ainonline.com/aviation-news/defense/2019-06-13/loyal-wingman-part-future-air-combat

Swanson, L., Jones, E., Riordan, B., Bruni, S., Schurr, N., Sullivan, S., & Lansey, J. (2012). Exploring Human Error in an RPA Target Detection Task. *Proceedings of the Human Factors and Ergonomics Society*, 328–332. https://doi.org/10.1177/1071181312561076

Taylor, R. M. (2006). Human Automation Integration for Supervisory Control of UAVs. *Virtual Media for Military Applications*, (12), 1–10. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA473313

Teigen, K. H. (1994). Yerkes-Dodson: A Law for all Seasons. *Theory & Psychology*, *4*(4), 525–547. https://doi.org/10.1177/0959354394044004

Tevithick, J. (2019). Air Force Wants Its XQ-58A Valkyrie Drone To Help F-22s And F-35s Talk To Each Other. Retrieved November 12, 2019, from The Drive website: https://www.thedrive.com/the-war-zone/30988/air-force-wants-its-xq-58a-valkyrie-drone-to-help-f-22s-and-f-35s-talk-to-each-other

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016). A Literature Review on the Levels of Automation During the Years. What are the Different Taxonomies that have been proposed? *Applied Ergonomics*, *53 Pt A*, 190–202. https://doi.org/10.1016/j.apergo.2015.09.013

Wassmuth, D., & Blair, D. (2018). Loyal Wingman, Flocking, and Swarming: New Models of Distributed Air Power. Retrieved November 12, 2019, from War on the Rocks website: https://warontherocks.com/2018/02/loyal-wingman-flocking-swarming-new-models-distributed-airpower/

Wickens, C. D. (1984). Processing Resources in Attention. In *Varieties of Attention* (pp. 63–101).

Wickens, C. D. (2002). Multiple Resources and Performance Prediction. *Theoretical Issues in*

Ergonomics Science, *3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wickens, C. D., Mavor, A. S., & McGee, J. P. (1998). *The Future of Air Traffic Control: Human Operators and Automation*. Retrieved from https://siris-libraries.si.edu/ipac20/ipac.jsp?&profile=liball&source=~!silibraries&uri=full=3100001~!628070~!0#focus%0D

Wiener, E. L. (1985). Beyond the Sterile Cockpit. *Human Factors*, *27*(1), 75–90. https://doi.org/10.1177/001872088502700107

Wiener, E. L. (1989). Reflections on Human Error: Matters of Life and Death. *Proceedings of the Human Factors Society 33rd Annual Meeting*, 1–7.

Wiener, E. L., & Curry, R. E. (1980). Flight Deck Automation: Promises and Problems. *Ergonomics*, *23*(10), 995–1011.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind Human Error: Cognitive Systems, Computers, and Hindsight*. Wright Patterson Air Force Base.

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rpaidity of Habit Formation. *Comparative Neurology and Psychology*, *18*(5), 459–482.

Young, M. S., & Stanton, N. A. (2002). Attention and Automation: New Perspectives on Mental Underload and Performance. *Theoretical Issues in Ergonomics Science*, *3*(2), 178–194. https://doi.org/10.1080/14639220210123789

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*
03/26/2020

2. REPORT TYPE
Graduate Research Paper

3. DATES COVERED *(From - To)*
September 2018-March 2020

4. TITLE AND SUBTITLE
Human Performance Modeling: Analysis of the Effects of Manned-Unmanned Teaming on Pilot Workload and Mission Performance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Andrews, Jinan M, 2d Lt (AFIT/ENV)
Meador, Douglas P, Dr (AFRL/RQVI)
Miller, Michael E, Dr (AFIT/ENV)
Rusnock, Christina F, Lt Col (AFLCMC/WWU)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-MS-20-M-185

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
711 Human Performance Wing
Timothy Webb, Dr.
852 Wright Ave Q,
Fairborn, OH 45324
Timothy.Webb@wpafb.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
An IMPRINT model was developed to predict a pilot's cognitive workload levels and mission performance in a simulated environment when interacting with the cockpit and multiple UAVs. This research concluded that peaks in workload occur for the pilot during periods of high communications load and this communication may be degraded or delayed during air-to-air engagements. Nonetheless, autonomous control of the UAVs through Vector Steering, Pilot Directed Engagements, and Tactical Battle Management would enable pilots to control up to 3 UAVs and their own aircraft against 4 enemy targets, while maintaining acceptable pilot cognitive workload in an air operation.

15. SUBJECT TERMS
Human-Agent Interactions, Unmanned Aerial Vehicles, Manned-Unmanned Teaming, Mental Workload, Improved Performance Research Integration Tool, Human Performance Modeling

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	141	Lt Col Christina F. Rusnock (AFLCMC/WWU) (321) 300-5231 Christina.Rusnock@us.af.mil

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.