The Pennsylvania State University

The Graduate School

MONITORING ACROSS HUMAN-AUTONOMY TEAMS IN ADVANCED AIR MOBILITY OPERATIONS

A Thesis in

Aerospace Engineering

by

Meghan Fickett

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2024

The thesis of Meghan Fickett was reviewed and approved by the following:

Amy Pritchett
Professor, Head of Aerospace Engineering
Thesis Advisor

Alan Wagner
Associate Professor of Aerospace Engineering

Eric Johnson
Professor of Aerospace Engineering

Jack Langelaan
Professor of Aerospace Engineering
Director of Graduate Program

ABSTRACT

Monitoring is a crucial element of human-autonomy teaming in novel aviation operations. Across a distributed team of agents, cross-checking or monitoring is needed to identify safety-critical situations. This thesis examines the impact that varying information distributions across the team, and when monitoring occurs, has on monitoring performance.

This thesis applied an agent-based simulation software, Work Models that Compute (WMC), to examine a Concept of Operations (ConOps) for Advanced Air Mobility (AAM). Within a day's operation involving several agents, 5 electric vehicles, and 30 missions totaling 60 flights, this case study introduced a degraded vehicle battery that would, if undetected, eventually result in the vehicle departing without sufficient energy to complete its flight.

The monitoring was varied in two ways. First, the algorithm used to assess the scenario was varied to reflect different potential information distributions across the team, in which the monitor would know the schedule, current state, requirements for upcoming flights, and/or predictions of state upon which the day's operations schedule is based. Second, the monitoring was conducted at different times: before, during, and after each flight.

Examining achievable true positive and false positive detection rates for each variation of monitoring, none of the types of monitoring were able to achieve perfect performance. Instead, monitoring accuracy in detecting a degraded battery varies significantly with the information distribution across the team and the timing of when the monitoring is conducted. Further, monitoring increases agent task load and the amount of information needing to be transferred between agents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This thesis would not have been possible without the significant help of my community of colleagues, mentors, family, and friends. Submitting this research has been the culmination of years of effort and was driven in large part by the incredible support of several individuals whom I wish to acknowledge here.

I wish to thank Dr. Amy Pritchett, my advisor, as well as Dr. Jack Langelaan, Dr. Eric Johnson, and Dr. Alan Wagner for serving as my thesis committee, offering their guidance and encouraging my scientific advancement.

Entering graduate school as a student with a non-engineering background at the height of the COVID-19 pandemic was incredibly daunting. Throughout my time at Penn State, Dr. Pritchett has been unendingly patient and supportive of me, introducing me to the world of human factors engineering and encouraging my unique path through this program as I began full-time employment in 2022 and worked to finish my degree remotely. Her expertise in this field and her immensely kind nature have been crucial to my success in this program, and I feel incredibly blessed to have had her as my advisor. I cannot thank her enough.

I wish to thank my colleagues in Dr. Pritchett's research group: Amber Villa, for her kindness and consistent help as I learned how to use WMC, and Ishan Ghimire, who offered his helpful perspective as I completed data analysis of my results. I also wish to thank the larger collaborative group of this project, including the following individuals: Dr. Cody Fleming from Iowa State University and Minghui Sun from University of Virginia, Natasha Neogi and Jon Holbrook from NASA Langley, and Dr. Martijn IJtsma from Ohio State University for hosting the WMC repository.

Starting a new degree program a state away from your family is difficult. Starting a program in the middle of COVID isolation is even more difficult. For this reason, I wish to thank

Dana Mikkelsen for being a great friend during this program, providing encouragement and creating a sense of normalcy through this extraordinary time. I also wish to thank my friends Chanelle Graham, Mariah Bowie, and Rachel Borczuch, for supporting me from near and far since childhood.

At times, the graduate school process has felt like a different world, with different rules. It’s easy to feel lost. For this reason, I want to thank my Aunt Debra and her colleague, Dr. Mohamed El-Aasser, for providing guidance as I navigated academia beyond my undergraduate degree.

I now wish to thank my family, for without them the challenges of this time would have felt insurmountable. To my mother, father, and brother, thank you for consistently reminding me that I am capable, strong, and loved. To my dog, Appa, thank you for never letting me go a day without chaos, joy, and laughter. And to my partner, Jim, thank you for your never-ending love; you support me in everything I do, and that has made this possible.

Lastly, I wish to acknowledge and thank my family that is no longer Earth-side, but from whom I continue to feel love, support, and spiritual guidance always. To my great-grandmother, Margaret Rose Krieg Seeley, thank you for inspiring me to pursue what will truly make me happiest. To my grandfather, Francis William Hartzell, thank you for giving me the confidence to solve any problem that comes my way. And to my grandmother, Margaret Seeley Hartzell, thank you for loving me, every day, forever.

This material is based upon work supported by the Aeronautics Research Mission Directorate’s System-Wide Safety project under NASA Grant No. 80NSSC19K1702. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of NASA.

Chapter 1

Introduction

Aviation operations increasingly envision applying autonomous capabilities as more than simply a tool that humans use, but as a genuine team member, capable of collaborating and equally participating in actions alongside human agents (Pritchett, Portman, & Nolan, 2018). However, the commercial aviation industry requires the highest levels of safety. One team interaction that contributes to safety is monitoring, in which one an agent is assigned authority to execute an action and a different agent is assigned responsibility for the outcome of the action and "monitors" the performance of the action (Villa, 2022). Unfortunately, there is little research as to how monitoring looks in practice within a distributed team of agents.

Past research has largely focused on aviation monitoring within the construct of a pilot monitoring autoflight systems, rather than exploring monitoring within a distributed team of agents both inside and outside of a vehicle (Mumaw, Billman, & Feary, 2020). In addition, the Concept of Operations (ConOps) for Advanced Air Mobility (AAM) published by NASA describes monitoring as a necessary element in a mission but does not elaborate on what information is required to monitor and when the monitoring should occur (Deloitte Consulting LLP, 2020). This thesis examines monitoring within a distributed human-autonomy (H-A) team in AAM operations using an agent-based simulation of a day's operation involving several agents collectively executing 30 cargo delivery missions with 5 electric vehicles. This case study introduces a degraded vehicle battery that, if left undetected, would eventually result in the vehicle departing with insufficient energy to complete its flight.

The objective of this thesis is to explore two significant aspects of monitoring, and the monitoring is varied in two ways. First, the monitoring action algorithm used to assess the scenario is varied to reflect different potential information distributions across the human-autonomy team, in which the monitoring agent would know the current vehicle state, requirements for upcoming flights, or predictions of the state upon which the day's operations schedule is based. Second, the monitoring is conducted at different times: before, during, or after each flight. Monitoring performance is assessed based on the observed frequencies of true positives and false positives that the monitoring agent achieves given when the action is conducted and the information upon which it is based. In addition, monitoring's effect on agent task load is assessed using action trace and information transfer calculations.

Chapter 2 of this thesis reviews H-A teaming overall, then Chapter 3 reviews literature regarding monitoring in aviation overall, and where monitoring is discussed in the relevant ConOps for AAM. Chapter 4 details the case study applied in this thesis, which follows the case study used in (Villa, 2022), and describes how it is modelled in the simulation framework "Work Models that Compute" (WMC). Chapter 5 defines several information distributions across the H-A team and the monitoring algorithms that each affords. Chapter 6 details the results of the simulation in terms of the observed frequencies of true positive and false positive detections achievable with the different timing of monitoring actions and the different information distributions the monitoring is based on. Chapter 7 concludes with a review of the implications of these results for monitoring within distributed H-A teams.

Chapter 2

A Review of Human-Autonomy Teaming

2.1 Human-Autonomy Teaming in Aviation

"Autonomy" is defined as the capability of machines to perform tasks that have been previously assigned to humans (Pritchett, Portman, & Nolan, 2018). With improvements in autonomous capabilities, there now exist additional opportunities for it to have a greater role in civil aviation. However, due to the high standards of safety for civil aviation, it must interact with humans in a way that supports safety. This extends beyond traditional views of a human supervising automation to "human-autonomy (H-A) teaming". A team is defined as "a group of two or more individuals who must interact cooperatively and adaptively in pursuit of shared valued objectives…team members have clearly defined differentiated roles and responsibilities, hold task-relevant knowledge, and are interdependent." (Cannon-Bowers, Salas, & Converse, 1993). For example, military operations such as those in the United States Air Force favor operational models of interdependent agents that comprise humans and flexible autonomy, with each agent being capable of taking initiative on some system elements, but also remaining coordinated with the larger system goals and situational awareness, and being capable of adapting to evolving dynamics (Zacharias, 2015). However, the civil aviation industry does not have a "rigorous, systematic, and cost-effective method" to demonstrate safety for approving H-A teaming technologies (Pritchett, Portman, & Nolan, 2018).

H-A teaming considerations, including team architecture, information flows, and activities, should be included in early design documents such as Concepts of Operation (ConOps). The design documents also need to allocate actions to the agents within the team.

However, research regarding this topic has historically focused on high-level descriptions of function allocations, and there are a number of details that are not addressed in current research, such as how system designers should account for cross-checking and monitoring within the H-A team (Pritchett, Portman, & Nolan, 2018).

Introducing an automated agent into a system is akin to adding an additional team member (Christoffersen & Woods, 2002). However, past studies have applied models of trust and social judgement to the concept of H-A interaction and have noted that automation developed at the time did not have the level of teamwork skills that humans naturally have (Bass & Pritchett, 2008; Hollnagel & Woods, 2006; Woods, 1985). For example, unlike human agents, automation cannot possess a sense of responsibility and therefore cannot be assigned responsibility for the outcomes of actions (Sarter & Woods, 1997). To be a good team member, agents must be able to both anticipate each other's information needs and deliver that information at beneficial times, as well as estimate the workload of each agent (Entin & Entin, 2001); however, automated agents tend to interrupt human agents due to their inability to sense whether the agents will benefit from the information being presented at that time (Christoffersen & Woods, 2002). For agents to correctly determine whether to interrupt another agent with information, they must be able to evaluate the importance of the information in question as well as the status of the agent they interrupt (Pritchett, Portman, & Nolan, 2018). Difficulties introducing automation into a team are often the result of an insufficient level of coordination between the human and automated agents (Christoffersen & Woods, 2002).

A beneficial team structure supports the "complementary team mental model", defined in (Sperling & Pritchett, 2011) as the state in which each member of the team has the information required to perform their own tasks, has the information regarding what other agents know, and knows what information of theirs will be needed by other agents and when that transfer must occur. Understanding the information that other agents possess and the information needs of other agents contributes to team situational awareness, a crucial element of a well-functioning team. However, current designs require the human agents to act as translators, reflecting the overall mission environment to the automation.

An additional concept for consideration is teamwork, i.e., additional actions created when agents need to coordinate and communicate with each other beyond their normal taskwork (Pritchett, Feigh, Kim, & Kannan, 2014), demonstrated in Figure 2-1.

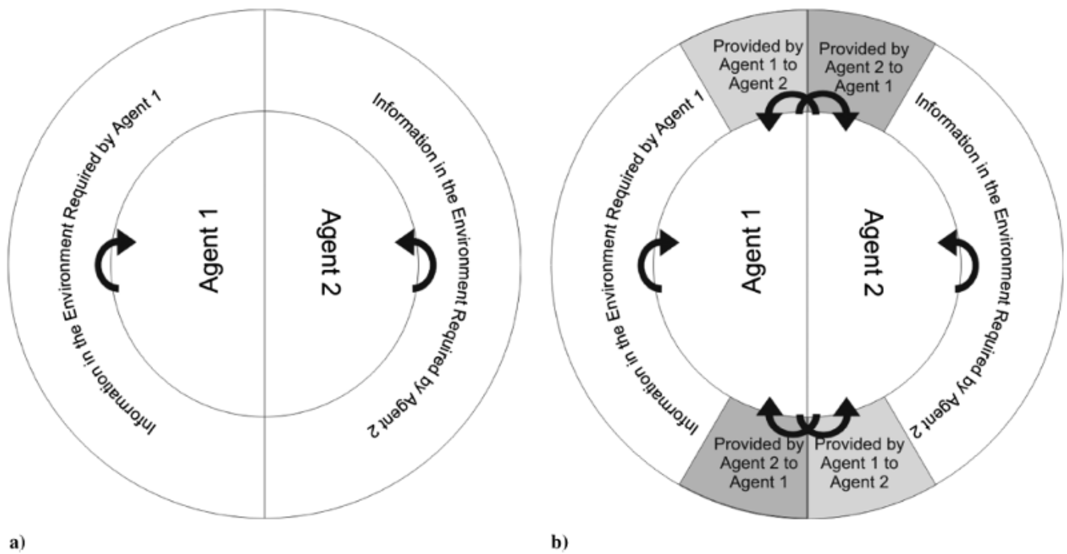

Figure **2-1**: Workload structure for a) taskwork only and b) taskwork and teamwork (Pritchett, Feigh, Kim, & Kannan, 2014).

The concept of agents functioning as a team is not new in the civil aviation industry. In fact, team structures have been a central element of civil aviation. Within aircraft, groundcrews, and air traffic management, teams of agents must function to complete actions and share resources to support a common mission. Within the National Airspace System (NAS), no single agent can perform all of the required functions for an aircraft. However, involving machines in a team structure is often difficult. In their review of H-A teaming for NASA Aeronautics, Pritchett, Portman, and Nolan conducted stakeholder interviews with multiple experts in the fields of human-machine interaction and aviation (2018). Several industry stakeholders stressed the high level of difficulty that exists when designing autonomous agents that are capable of handling unexpected events such as those that exist in aviation. Regarding automation in civil aviation, a robotics and human-robotic interaction researcher interviewed stated that “successful autonomy has not been demonstrated in these environments.” (Pritchett, Portman, & Nolan, 2018). Likewise, a senior government human factors scientist discussed humans in teams, stating that “the human is the glue that holds the system together especially when things go wrong, the human brings safety”. In that same vein, a senior government safety researcher stated that “somehow we need to be able to have the human operator able to act to assure overall system safety.” (Pritchett, Portman, & Nolan, 2018). In addition, autonomy performing properly often depends on specific conditions being met, limiting its ability to be applied to complex situations. This quality of autonomy is called “brittleness” and presents significant difficulties when attempting to implement automation into dynamic and/or uncertain environments.

This brittleness, and the legal designation of responsibility as having to ultimately fall to a human, fundamentally limits the current H-A teaming architecture. Current architectures tend to follow supervisory control structures, in which the automated agents assume functions that are well structured and adhere to expected environmental conditions, and the human agent supervises the automated agent. This supervision creates additional cognitive workload for the human agent. The current state of the of the art in the aviation industry is that “pilots mitigate safety and operational risks on a frequent basis, and the aviation system is designed to rely on that mitigation.” (PARC/CAST, 2013). In the stakeholder interviews performed by Pritchett, Portman, and Nolan, a senior government safety researcher stated that “if you automate a system to cut crew costs, [the] final new system must have at least the same level of safety, it must not increase the opportunity for errors or catastrophic events” (2018). Therefore, it is important to deliberately design the H-A team architecture from the beginning of system development, such as when developing new ConOps. However, additional research is required to identify H-A teaming architectures that support safety, including cross-checking and monitoring, and timely interventions and adaptations.

Automation can be involved in monitoring efforts, but human agents must retain the highest level of responsibility. As codified in Federal Aviation Regulation §91.3, “The pilot in command of an aircraft is directly responsible for, and is the final authority as to, the operation of that aircraft”, and “In an in-flight emergency requiring immediate action, the pilot in command may deviate from any rule of this part to the extent required to meet that emergency” (Pritchett, Portman, & Nolan, 2018). This legal responsibility exists even with the introduction of new automated technologies. With Unmanned Aircraft Systems (UAS), a responsible human agent must be designated within the system, who possesses the necessary information and control authority to properly maintain safety, even if automation is executing a significant portion of the system functions.

> Automation technology often behaves incorrectly due to issues with incorrect assumptions or operating outside their boundary conditions. Thus, humans’ monitoring of automation needs to examine not only for failures within the technology, but also system conditions that make the automation’s functioning-according-to-specifications contextually correct.

2.2 Allocation of Authority and Responsibility in Human-Autonomy Teams

Beyond the structure of an H-A team, Feigh and Pritchett proposed the following five key requirements for function allocation to humans and autonomous agents (Feigh & Pritchett, 2014). It should be noted that while Feigh and Pritchett discuss the allocation of system functions, this thesis focuses on the allocation of actions to agents; these requirements still apply.

1. Agents are to be assigned functions that they are capable of performing:

H-A function allocation often focuses on authority assignments, defining the agent who is assigned the execution of the function in an operational sense. When automation is capable of initiating or conducting its own actions, it can be considered an agent (Lee J. D., 2006). When automation has the possibility to be placed beyond its boundary conditions in its environment, and mismatches between authority and responsibility exist within the function allocation, humans are required to monitor the automation (Bainbridge, 1983; Wiener & Curry, 1980). However, humans have consistently been found to be ineffective at monitoring, especially when performing other non-monitoring tasks (Lee & Moray, 1992; Molloy & Parasuraman, 1996). Therefore, it is also important to consider the assignments of responsibility as well.

2. Each agent must be able to perform its entire set of assigned functions:

Under realistic operation conditions, each agent should be capable of performing its entire task load assignment, including functions it is assigned authority or responsibility for. If an agent is assigned functions that overload its total task load capabilities, it will be unable to perform. It is important to not only examine the total number of functions assigned to an agent, but also the distribution of task load assigned through time. Automation can help reduce the average workload by reducing the manual execution functions that human agents must perform. However, past operational research has also demonstrated that at times, automation can create a workload spike for human team members (Bainbridge, 1983; Billings, 1997; Wiener, 1985). These workload spikes may be caused by cognitive functions assigned to the human agent consisting of monitoring and intervening with automated functions and can create significant information and decision-making requirements for the human. It is therefore important to consider metrics of information and task load when assessing a function allocation (Pritchett, Portman, & Nolan, 2018). In addition, any assessments of the total workload of a human agent should account for the monitoring functions that have been either explicitly or implicitly assigned to them.

3. The function allocation must be realizable with reasonable teamwork:

When in a team environment, whether that team is entirely made of human agents or contains automated agents, any allocation of taskwork functions will create additional teamwork functions that coordinate taskwork within the team. Teamwork actions can include human-to-human agent communication, human-automation interaction, and coordinating the timing and conduct of taskwork. Each different function allocation creates its own set of teamwork functions. The impact of those added functions must be assessed based on the requirements that the agent can perform each function it is assigned as well as the total collection of all of its assigned functions. Likewise, when automation is viewed as a team member, communication between human and automated agents must be investigated. Team members, whether they be human or automated, must be able to anticipate other agents' information needs and present that information at useful times that do not create undue interruptions (Entin & Entin, 2001; Hollenbeck, et al., 1995; Hutchins, 1996). Therefore, the timing of information requirements and information transfer must be considered when designing function allocations.

4. The function allocation must support the dynamics of the work:

In team environments completing dynamic work missions, many activities are interdependent and therefore must be considered in conjunction with one another. When designing a function allocation, it is important to determine if function assignments will require coordination between agents, cause agents to wait on or interrupt each other, or create situations where workload may significantly accumulate. These function assignments should support the dynamics of the work environment and the interactions between agents. This also includes understanding the difference in behavior between human and automated agents. While humans can adapt their overall behavior to maintain strength and resilience in a system during off-nominal situations, automated agents are typically brittle during off-nominal situations. A human agent's ability to adapt to the situation can be limited by an overly rigid function allocation, especially when human-automation interaction requires a specific sequence of activities from the human agent (Feigh, 2011). In the case of monitoring, dictating that humans are engaged in monitoring actions in some capacity may limit their ability to participate in other activities during that time, and limit their ability to adapt to other circumstances.

5. The function allocation should be the result of deliberate design:

Function allocation is the result of deliberate design choices that must be made early in the design of documents such as a ConOps (Dearden, Harrison, & Wright, 2000). The National Research Council has requested that system designers use a more systematic approach to evaluating and designing concepts of operation and function allocations in aviation (NRC, 2003). Function allocations that are created early in the design process can be a cost-effective method for determining technology specifications and identifying the expected information and interaction requirements between agents. Developing ConOps to consider allocations of authority and responsibility and their effect on monitoring and information distribution between agents provides crucial design assistance.

Chapter 3

A Review of Monitoring Literature

In their review of H-A teaming, Pritchett, Portman, and Nolan note the following regarding monitoring: “When a human operator is directly and continuously controlling a process, their monitoring behaviors are naturally part of their control activities. As the human supervisor instead transitions to monitoring an automatous function through inputs at discrete points of time, s/he instead needs to anticipate how the machine will function, and the impact of this function, over the period of time up to their next input. As the time between human inputs increases, this anticipation increasingly depends on an accurate model not just of the controlled process, but also of the automation and of the operating context” (Pritchett, Portman, & Nolan, 2018).

ConOps may require monitoring between agents. For example, the NASA ConOps for AAM discusses the relationship between agents in several Unmanned Aviation scenarios. For example, the air traffic control (ATC) operators have “access to all available real-time information about UAM operations, but does not generally monitor UAM operations; rather, ATC is alerted only in the case of an emergency or when UAM operations depart from their desired parameters.” (Deloitte Consulting LLP, 2020). The document states that a network of agents exists called PSUs, or Providers of Service to UAMs (UAM being an out-of-date term, now replaced with AAM), to provide air traffic management. Information that enables safe flight operations and requires constant updating (for example, aircraft location and proximity to obstacles) is processed on the UAS and any communication that is sent between vehicles is sent to the PSU for status monitoring. The PSU network can be accessed by public safety organizations to monitor the UAS operating environment. UAS operators can also monitor the

PSU network and share their takeoff and landing information with a PSU to then distribute that through the network.

The NASA ConOps also describes the relationship between automated and human agents to distribute information and monitor operations. Before a UAS takeoff, the automated systems on the vehicle continually assess the vehicle's actual status and compare that to its projected status and operations plan. The automated systems alert fleet operators if a noticeable difference is detected between the actual and expected conditions. In the same manner, the human fleet operator will monitor and maintain the operations plan of the vehicle fleet alongside the PSU and notifies the vehicle automated systems if any updates are needed to the operations plan. In addition, the vehicle position and separation from other vehicles is monitored throughout the departure sequence. During the flight, the vehicle's automated systems and crew then monitor the current flight position relative to the predicted position according to the operations plan. Significant differences are again communicated to the fleet operator.

In the ConOps, monitoring is listed as a responsibility for multiple agents, both automated and human. The PSU agent, which is automated, is assigned the responsibility of supporting "operational planning, aircraft deconfliction, conformance monitoring, and emergency information dissemination." The "aircraft crew" agent, either on or off-board, is assigned the responsibility of providing communication and limited loop monitoring of vehicle health and status to be reported to the fleet operator. The fleet operator is assigned monitoring throughout the vehicle flight, monitoring the vehicle and its conformance to the operations plan.

Monitoring is also assigned to agents in the event of off-nominal situations. In the event of an emergency, the person in charge of the vehicle must acknowledge the critical failure and monitor the initiated actions taken by the vehicle operator agent(s). The person in charge may also initiate additional actions beyond monitoring to ensure vehicle safety.

However, the ConOps does not outline the specific types of information that are provided to each agent to facilitate monitoring, and at what times the information should be delivered. These are crucial details that affect how the monitoring will occur in real-life situations, as well as the monitoring agents' ability to detect and prevent safety-critical situations. Villa applied this ConOps in an investigation of task-limited agents in distributed team operations (Villa, 2022). She found that the lack of specification regarding monitoring in the ConOps fails to address potential negative impacts that monitoring may have on increased agent task load and information transfer requirements.

Unfortunately, there are not many investigations of monitoring in literature. The most substantive example is of an analysis of pilots' monitoring of autoflight systems (Mumaw, Billman, & Feary, 2020). This paper investigates monitoring in practice between two (human) pilots; a flying pilot (PF) and monitoring pilot (PM). In current-day operations, the PM is expected to do the following actions:

- Monitor current and projected flight path and energy of aircraft at all times
- Support PF at all times
- Monitor aircraft state and system status

In addition, both pilots are engaged in these related activities:

- Planning for managing operational tasks
- Possessing operational knowledge and understanding to determine what information is relevant, how attention is allocated, and how the information is comprehended
- Switching attention to be responsive to emergent events and ensure flight path targets are being met
- Communicating a shared understanding of Flight Path Management (FPM) objectives to coordinate monitoring activities

- Identifying deviations from the expected state to ensure that they are called out and managed

However, evidence from past research has demonstrated that humans can have difficulty monitoring automation (Parasuraman, Sheridan, & Wickens, 2000; Parasuraman & Riley, 1997), and their ability to maintain a mental awareness of information in the environment is diminished when they passively monitor the decisions of an automated agent (Kaber, Omal, & Endsley, 1999).

Failures of monitoring specifically in FPM imply that the PM has failed to be aware of and understand the vehicle's state relative to the flight path and cannot note the deviations from the operations plan (Mumaw, Billman, & Feary, 2020). An example of monitoring failures in flight is the Turkish Flight 1951 in 2009, in which the flight crew was unaware that the vehicle's airspeed had decreased significantly. The degradation in speed resulted in a crash with nine fatalities. During the Ethiopian 409 crash in 2010, the flight crew was unaware of a vehicle slow down and pitch change for 27 seconds until the stick shaker came on. Flight degradations resulted in a crash with 90 fatalities. Loss of Control (LOC) incidents such as these that involved monitoring failures were investigated by the Commercial Aviation Safety Team's (CAST) Airplane State Awareness (ASA) team. CAST found that Loss of Airplane State Awareness was a significant factor in at least half of all Loss of Control Inflight (LOC-I) events in commercial aviation (CAST, 2014).

There are several failure categories that Mumaw et al explore in their 2020 paper (Mumaw, Billman, & Feary, 2020). These categories include:

- Imminent upset (CAST 2014). While alerting systems are in place for many conditions, in multiple LOC accidents, there was either no alert for the basic flight path hazard, or the alert indication failed to get the pilot's attention (Mumaw, Haworth, & Feary, 2019)

- Failing to meet flight path targets
- Incorrect airplane state or configuration
- Failing to identify important changes in the flight environment
- Failing to maintain awareness of crew resources
- Failing to call out a deviation from the FPM target
- Failing to intervene when a deviation is not being managed

While Mumaw et. al state that a human agent will be the PM in their investigation, there are human limitations that reduce the effectiveness of FPM. These include:

- High workload
- Interruptions or distractions
- Vigilance/awareness failures, such as those caused by periods of low or high workload (Lysaght, et al., 1989; Miller & Parasuraman, 2007)
- Mind-wandering
- Stress and fatigue
- Channelized attention
- Inattention blindness
- System automation

Throughout the course of a mission, the monitoring agent will often be managing several other tasks, and human attention is a limited resource. With the introduction of monitoring, agent workload can increase quickly and there may be moments when the monitoring agent cannot dedicate any energy/attention to monitoring (Mumaw, Billman, & Feary, 2020). Therefore, when assessing how monitoring will exist within a system, one must understand the overall effect on attention resources and task load experienced by agents when monitoring is introduced.

To have an effective monitoring model, system designers require two elements: a situation model, and a sensemaking cycle (Mumaw, Billman, & Feary, 2020). A situation model represents a pilot's understanding of the current situation and can be based on different levels of detail and accuracy. Three types of input may be used to develop the situation model (Hutchins, Holland, & Norman, 1985), such as:

- Data and information from the world acquired from monitoring activities
- Relevant mental models, which are a collection of facts and rules that help a pilot understand the typical behavior of a system
- Knowledge the pilot obtains through operational experience

It is important to understand that while automated agents may be able to possess some of these inputs, automated agents cannot create their own situational models that benefit from their own knowledge or operational experience. In addition, the rigidity of automated tools can limit a team's ability to adapt to off-nominal situations (IJtsma, 2019). Therefore, a combination of human-automation monitoring may be beneficial in this case, where automation can serve as a tool to help the human operator, who then functions as the final monitor. Situation models are also affected by the accuracy of the information that makes up the model: The information resources, and the timing at which that information is presented to agents, can affect the situation model from which agents derive monitoring decisions.

The sensemaking cycle describes the theory that monitoring is tied to the process of understanding the ongoing environment and situation. Throughout flight missions, resources and variables are constantly changing, and having a mental model can alert the pilot to the important variables that change frequently and must be tracked. Monitoring operates in a cycle that is guided by, and updates, the situation model. There are three key processes to this cycle:

- Identify gaps in understanding
- Gather relevant data/information
- Identify appropriate actions

In a broader scope, monitoring is supported by effective communication between all agents that helps maintain a shared understanding of the environment, potential threats, and mission objectives. Monitoring is also enabled by intervention, in which the assigned agent effectively ensures that FPM targets are achieved and deviations from the norm are addressed/rectified (Mumaw, Billman, & Feary, 2020). While automated alerts may identify an issue or deviation, the monitoring agent (that is likely human) will need to ensure that correct action is taken to rectify the issue. For automated agents to complete all monitoring requirements, they would need to be designed with the capabilities needed to identify issues, decide what action is required, and perform the action necessary to rectify the issue.

Past literature developed mathematical models of monitoring (Carbonell, 1966), updating models over time and applying them to aviation (Wickens, Goh, Helleberg, Horrey, & Talleur, 2003). These papers developed a model of a human monitoring agent whose prime responsibility is monitoring a small set of indicators for deviations.

Mumaw et al (2019) also reviewed two types of training approaches to monitoring: the “bottom up” training method, which focuses on where the pilots look during the mission, and the “top down” training method, which focuses on integrating and understanding information. For each training approach, there has been research addressing the measurement of non-technical behavioral and verbal indicators to estimate agent behavior within a team (Flin, et al., 2003).

Several papers focused specifically on the "bottom up" training method of monitoring (Bellenkes, 1999; Bellenkes, Wickens, & Kramer, 1997). These papers describe several attempts that have been made to characterize where experienced pilots look during a flight so less-experienced pilots can learn proper eye-scanning patterns. They found that more experienced pilots tend to have shorter fixation periods on a single indicator with more frequent shifts across the interface and are more likely to sample secondary performance indicators to cross-check information. In 2018, Froger et al worked to train pilot gaze distribution at a general level, aiming to reduce pilot dwell times on a single area by training them to monitor their time spent on a task (Froger, Blättler, Dubois, Camachon, & Bonnardel, 2018).

Regarding the "top down" method of training, Mumaw et al found research that covered different types of research methods. These include:

- Embedding situational awareness training within the broader training of non-technical skills (Salas, Burke, Bowers, & Wilson, 2001; Holt, Boehm-Davis, & Hansberger, 2001).
- Focusing on training situational awareness specifically (Potter, 2001; Bolstad, Endsley, Costello, & Howell, 2010; Kearns, 2011; Bolstad, Endsley, Howell, & Costello, 2002; Ferrari, Spillmann, Knecht, Bektas, & Muehlethaler, 2018).

Mumaw et al found that there were only a few strong studies that assessed effectiveness of methods used to train pilots to monitor and assess situations (Holt, Boehm-Davis, & Hansberger, 2001; Hormann, Banbury, Dudfield, Lodge, & Soll, 2004). However, they noted that these may not reflect current pilot demographics and modern aircraft due to the studies' age.

Within the ATC domain, the controller's job lies in maintaining situational awareness and monitoring aircraft for potential separation violations or re-routing requirements. Multiple articles explore this domain (Knecht, Muehlethaler, & Elfering, 2016; Billinghurst et al., 2011;

Vu, Kiken, Chiappe, Strybel, & Battiste, 2013; Malakis, Kontogiannis, & Psaros, 2014; Hauland, 2008). However, Mumaw et al note that there are few studies specifically regarding training ATC skills.

Mumaw et al note that their literature review finds cross-domain themes between aviation and ATC that can be employed for training pilot monitoring (Mumaw, Billman, & Feary, 2020). They found that the target of training should be important for the work, be undeveloped in the trainees, and be a trainable target. In addition, system designers need to be able to identify the relevant information to monitor. Training pilots' understanding of how the systems are expected to operate allows them to build a long-term mental model, which will also benefit monitoring. Their findings included several failures that hinder monitoring performance, as well as the knowledge and skills that shape monitoring performance. Regarding training monitoring, system designers could train agents in either task/workload management or developing a situation model.

In addition, Mumaw et al outline a short history of institutional findings and training recommendations on pilot monitoring. Over several decades, there have been multiple aviation and safety organizations that have recognized concerns with pilot monitoring and awareness issues (Mumaw, Billman, & Feary, 2020). A 1994 study of 37 flightcrew-involved accidents between 1978 and 1990 identified 302 pilot errors, 23% of which were due to monitoring errors. Monitoring failures occurred in 31 of the 37 accidents. 23 of those accidents resulted in fatalities. (NTSB, 1994). In 2010, CAST formed a team to analyze LOC incidents and identified that failures in monitoring were considered to be significant contributors to these incidents (CAST, 2014). To address these issues, there have been several articles aimed at improving pilot monitoring performance (Federal Aviation Administration, 1996; PARC/CAST, 2013). These studies demonstrate the safety-critical nature of monitoring; if it is eliminated or performed incorrectly, accidents can occur with dire consequences.

Overall, Mumaw et al identified and organized past literature to demonstrate the importance of proper monitoring behavior within aviation and air traffic management and outlined common failure types for monitoring flight path management. They emphasize that proper task management, situation awareness, and a proper mental model are crucial for effective monitoring (Mumaw, Billman, & Feary, 2020). However, Mumaw et al focused primarily on FPM and monitoring within a 2-person team, with one human agent performing all monitoring of flight-path automation and one human agent performing all taskwork actions.

Chapter 4

Overview of WMC and Case Study

This thesis examines monitoring and information distribution across a team using the "Work Models that Compute" (WMC) simulation framework. Previous researchers have applied WMC to represent H-A teaming behavior in both aviation and spaceflight missions (Villa, 2022; Ma, IJtsma, Feigh, Paladugu, & Pritchett, 2018; IJtsma, Ma, Pritchett, & Feigh, 2019; IJtsma, 2019). Pritchett et al detail the WMC simulation structure in their paper "Work Models that Compute to Describe Multiagent Concepts of Operation: Part 1" (2014).

4.1 Work Models that Compute

The WMC simulation framework is a publicly available software created using C++ to support the design and analysis of ConOps. These ConOps specify the actions that need to be performed within a team of agents comprised of humans and automated agents. ConOps also specify either the function allocation or the allocation of authority and responsibility for each action to the agents. To complete actions correctly, agents need to view and act upon the entire collective work environment.

WMC work models comprise resource variables and actions. Each action contains a method that represents the work they execute by sampling the environment by getting resources and changing the environment by setting resources. Each action also contains a method, "next update time", that determines when the action will next need to be executed. This allows for either frequent updates of a continuously evolving physical state, such as guidance towards waypoints, or more episodic discrete actions, such as decisions to approve vehicle take-off. In

WMC, actions are modeled in detail outside of agent models. Therefore, the agent models do not contain any representation of the work, and teamwork and taskwork are represented by actions and resources.

When a WMC simulation is started, allocations of authority and responsibility are designated by creating links between actions to agents and resources, and reverse links are created between agents and resources to actions. The system designer does not need to predefine the links, as the simulation creates these automatically. Therefore, the system designer has more flexibility to add or replace different models of actions and resources to reconfigure work models and allocations of authority and responsibility without needing to alter agent models.

The simulation engine then begins by scanning the work model for actions and loading them into an action list. The engine then orders the actions based on their next update time, with the earliest update time being placed at the beginning of the list, and so forth. The simulation starts with executing the action at the top of the list, and after that action is performed the simulation clock is advanced to the action's update time. The action that was performed then declares a new update time and is sorted back into the action list accordingly. The next action at the top of the list is then executed. This process continues until the simulation is completed.

WMC is unique in its approach to agent models, as agent models are only brought forth by the simulation engine when the next action to be executed from the action list declares which agent has authority to execute it. The action is then passed to that authorized agent. If a "perfect" agent model is being used, the agent will immediately perform the action. If a system designer wishes to model specific attributes of agent behavior, those attributes can be included in a more extensive agent model that can then be applied during simulations of any work model. For example, agent models that contain a task load limit will delay actions assigned to them until they can be performed based on the limit, such as in (Villa, 2022). Agent models can therefore provide task load tracking, understand how an agent can manage multiple actions assigned to them, and determine whether an agent has access to the correct information and resources required to perform the action.

A full simulation work model within WMC can contain thousands of actions and resources. WMC's freedom to edit any aspect of the work model provides opportunities for testing different scenarios. For example, if a simulation designer is unsure what taskwork will be required in a ConOps, WMC allows for dynamically simulating the collective work and work environment to determine which actions and resources are required. In this thesis, WMC is used to determine what effect adding monitoring actions will have on the team's collective work and performance. This can enable the allocation of authority and responsibility to be further defined in future revisions of the relevant ConOps.

4.2 AAM Case Study

This thesis applies a ConOps written for NASA Aeronautics (Deloitte Consulting LLP, 2020). This is a new document and was not in its final draft stage at the time of this work. Villa's 2022 thesis first created a WMC simulation of this ConOps (Villa, 2022). This thesis will use the same case study as Villa and will instead explore monitoring performance and the effects of monitoring on information distribution and team behavior.

The ConOps in the WMC framework has three types of work models: the Command, Mission, and Vehicle. The actions within each work model are shown in Table 4-1.

Table **4-1**: WMC work models and their actions for the AAM ConOps (Villa, 2022).

Command Work Model	**Mission Work Model**	**Vehicle Work Model**
Assign mission	Request performance authorization	Assess battery
Approve performance authorization	Flight plan	Charge battery
Authorize flight	File operations plan	Systems check
Initiate takeoff planning	Receive operations plan approval	Takeoff
Allocate takeoff pad	Request flight authorization	Flight dynamics
Initiate departure sequence	Board/process payload	Manage waypoint progress
Approve takeoff	Request landing clearance	Maintenance
Issue landing clearance	Unload aircraft	
Allocate landing pad		
Initiate arrival sequence		
Confirm clear for landing		

The allocation of authority and responsibility is defined at the start of each simulation run. This case study uses the allocation shown in Table 4-2, which is the same as previously used by Villa (2022).

In WMC, monitoring actions are triggered when there is a mismatch between the allocation of authority and responsibility for the action next to execute: as the taskwork action is passed to the agent authorized to execute it, a parallel monitoring action is passed to the agent responsible for the taskwork action's outcome. Villa notes that monitoring resulting from a particular set of authority/responsibility (A/R) mismatches within an allocation of actions will affect the overall task load among the agents of a distributed H-A team. Not only will this affect the task load for the particular monitoring agent, it will also increase the total number of actions performed within the overall simulation, resulting in increases in information transfers and system performance resources. In Villa's allocation, responsibility, and therefore monitoring, is assigned as follows: the Mission Operator agent monitors all Vehicle Operator actions, and the Vehicle Operator monitors mission actions that are related to the vehicle. The Command agent will monitor any vertiport operator actions with A/R mismatches. The Vertiport Operator agent is not listed as responsible for any actions and therefore does not participate in any monitoring actions (Villa, 2022).

Table **4-2**: Example of allocation of authority and responsibility for the AAM ConOps (Villa, 2022).

Action	Authoritative / Executing Agent	Responsible / Monitoring Agent
Approve Taxi Takeoff	VertiportOp	Command
Allocate Takeoff Pad	VertiportOp	Command
Request Flight Authorization	MissionOp	VehicleOp
Takeoff	VehicleOp	MissionOp
Flight Dynamics	VehicleOp	MissionOp
Maintenance	VehicleOp	MissionOp
Charge Battery	VehicleOp	MissionOp

Different types of monitoring can be invoked within WMC: basic monitoring is modeled as only “getting” the resource(s) that the taskwork action “sets”, representing a simple check of the outcome of the taskwork; full monitoring is modeled as “getting” the resources that the taskwork action both “gets” and “sets”, representing some degree of duplicate assessment of what the outcome ought to be; extended monitoring allows for specific algorithms to be applied for monitoring any taskwork action. This thesis focuses primarily on extended monitoring actions.

This case study focuses on allocating responsibility to three agents: Command, Mission Operator (MissionOp), and Vehicle Operator (VehicleOp). The Vertiport Operator (VertiportOp) agent does not have responsibility for any actions in this case study. For some actions, a fifth default vehicle agent serves as only the authoritative agent.

The case study simulates a cookie delivery operation in State College, Pennsylvania. Five UAS make up the vehicle fleet, and each cookie delivery payload is 1.5 pounds. Collectively the vehicles conduct a total of 30 missions. Each mission consists of two flights: a mission flight to deliver the cookie payload to the intended delivery location, and a return flight to arrive back at the home vertiport located on the roof of the Hammond Building on the Penn State campus. Vehicle batteries can be charged following return flights, after a mission is completed. The WMC simulation used in this thesis included good-fidelity models of vehicle dynamics during a flight. The trajectory and flight profile of each UAS was modeled based on continuous simulation, at 50 Hz, of the vehicle acceleration, velocity, and attitude, and on models of the rotors and motors providing thrust and lift sufficient to capture the power required by and energy drain on the batteries.

This thesis examines two scenarios within this case study: a nominal case, where each vehicle battery performs as expected, and a degraded battery case, where each vehicle battery burns energy faster in flight, and charges slower, than the rates assumed by the actions that schedule flight times and battery charging periods.

Figures 4-1 and 4-2 show the timeline of vehicle battery level relative to “0 kW-s”, the lowest safety level, for a nominal case and degraded battery case, respectively. The blue line in the figures represents the battery level in kW-s over time for a full simulation run. The red “x” markings represent the predicted battery level, based on the process model of battery performance discussed below, at the end of each mission. The red horizontal line represents the “zero” level of the battery. It is considered a safety-critical event when the battery crosses below the zero line.

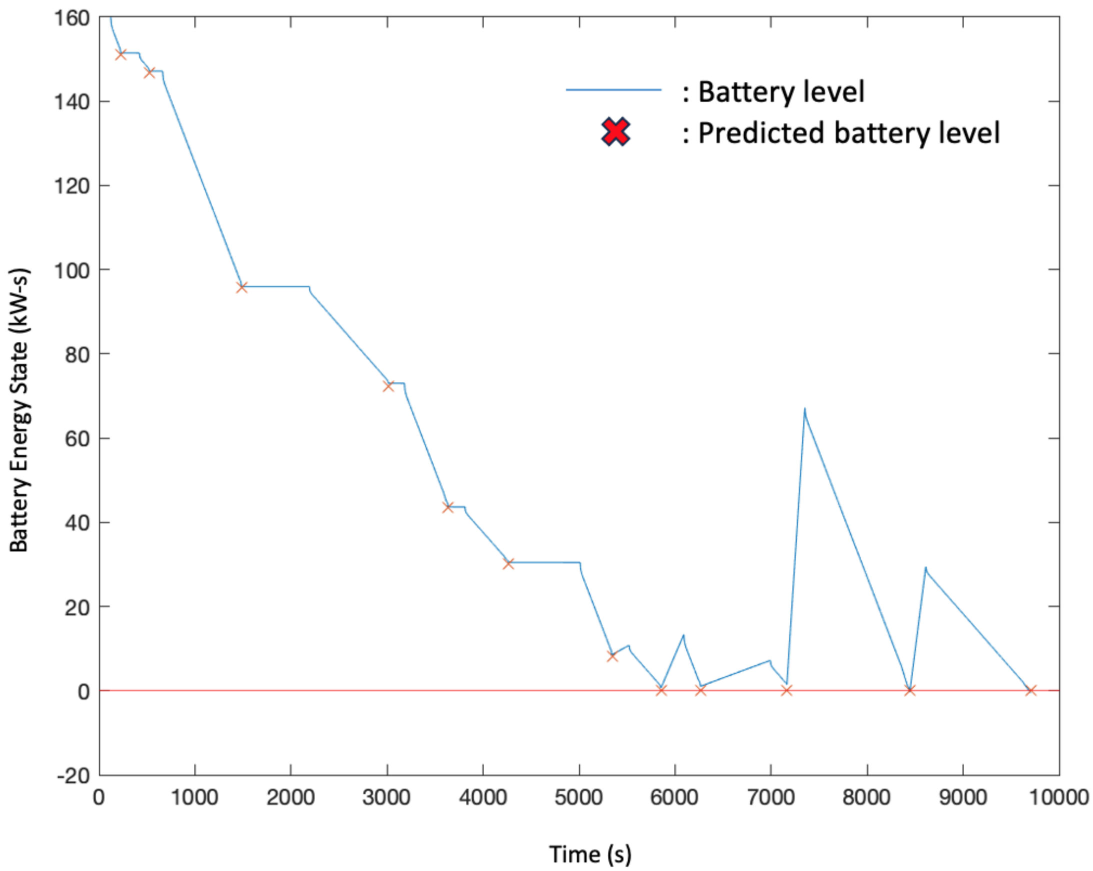

Figure **4-1**: Timeline of battery energy storage for one UAS with a nominally functioning battery throughout its flights and battery charge cycles.

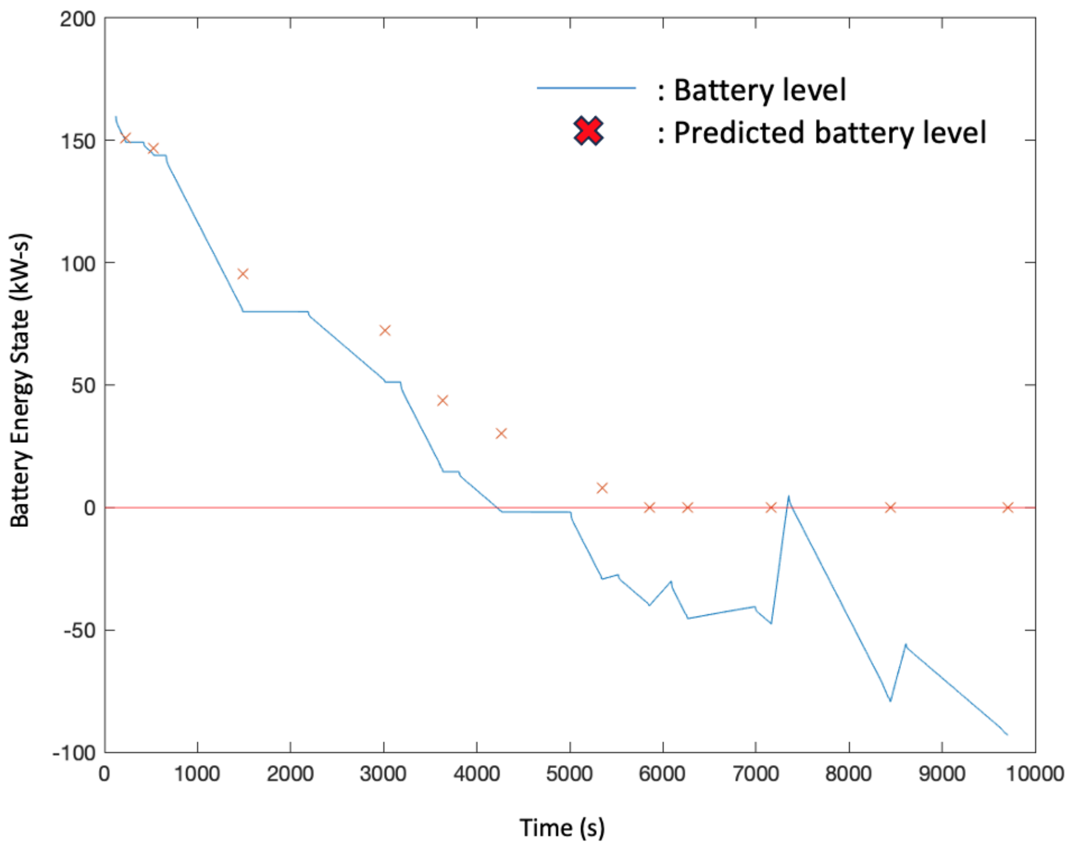

Figure **4-2**: Timeline of battery energy storage for one UAS with a degraded battery throughout its flights and battery charge cycles.

While the nominal battery never drops below the zero line because the schedule includes appropriate charging periods, the degraded battery can fall below the zero line, representing a safety-critical event. In addition, the nominal battery meets the predicted level at every checkpoint, while the degraded battery falls below the predicted level.

The battery model, assumed in the actions that plan the schedule of flights and charging times to keep battery levels in a safe range, approximates energy level with the following equations, created by Villa (2022):

$$Required\ Ascent\ Energy\ (kW \cdot s) = 0.0247 * (payload)^2 + 0.5504 * (payload) + 1.3$$

$$Required\ Descent\ Energy\ (kW \cdot s) = 0.0439 * (payload)^2 + 0.3765 * (payload) + 0.8$$

$$\frac{Required\ Cruise\ Energy\ (kW \cdot s)}{nm} = 0.3213 * (payload)^2 + 2.2153 * (payload) + 3.25$$

$$Mission\ Energy\ Requirements\ (kW \cdot s) = Ascent\ Energy + Descent\ Energy + \frac{Cruise\ Energy}{nm} * (Distance)$$

While nominal batteries will follow the battery model equations as written, degraded batteries burn energy at a 25% higher rate and they charge at (100/125) % of the nominal charging rate. If left undetected, the degradation of the battery will create safety-critical situations. Monitoring actions throughout the simulation create opportunities for identifying unsafe vehicle conditions before a safety-critical event occurs. Therefore, this thesis investigates the potential for monitoring to identify safety-critical behaviors, using a degraded battery as a case study.

Chapter 5

Modeling How Monitoring is Performed Based on Information Distribution

This thesis explores different algorithms of how monitoring may be performed. Each algorithm requires different distributions of real-time information, as well as different knowledge about the process being monitored.

This thesis uses, as a case study, monitoring sufficient to detect a degraded vehicle battery. As noted in the previous chapter, the schedule of flight times and charging periods is generated at the start of the day's operations based on a model of a nominally performing battery. Once the operation starts, monitoring is expected to occur across the team according to the information available to the monitoring agent. This chapter details the variations of information distribution across the team and the associated monitoring algorithm that each information distribution enables.

Monitoring Based on Schedule:

In this case, the responsible, i.e., monitoring, agent has access to the current plan of flight times and can confirm that flights are progressing according to this schedule. Monitoring examines the difference in time, $\Delta t,$ between the current flight time and schedule. If Δt reaches a certain level, indicating that the current flight time has deviated from the schedule beyond a certain point, an alert would trigger. This monitoring style mimics the real-life situation of a monitoring agent solely focused on schedule conformance.

Monitoring Based on Current State:

In this case, the responsible agent has access to the current state, represented in this case study by the actual battery level at the time of the monitoring action. The responsible agent does not, however, have access to the plan for upcoming flights and does not have the knowledge to assess whether the battery state is sufficient for them. In the event of a degraded battery, for example, the monitoring would trigger an alert if the battery percentage level for the vehicle falls below a certain setpoint.

Monitoring Based on Current State and Upcoming Requirements:

In this case, the responsible agent has access to the current state of the vehicle, in the form of the battery level and vehicle status, as well as the information required to determine the energy requirements of the upcoming mission. Monitoring will compare the actual battery level and required energy for the remainder of the upcoming mission, assuming a nominally performing battery.

Monitoring Based on Current State Compared to Predictions

In this case, the responsible agent has access to the current state of the vehicle, in the form of the battery level and vehicle status, as well as the battery level predicted earlier, when the schedule of flights and charging intervals was made. Therefore, monitoring will trigger based on the difference between the actual battery level and the predicted battery level.

Table 5-1 shows these information distributions and the monitoring associated with each. This thesis collects data for the "Current State", "Current State & Upcoming Requirements", and "Current State & Predictions" information distributions.

Table **5-1**: Information distribution and associated monitoring.

	Schedule	**Current State**	**Upcoming Requirements**	**Predictions**
"Schedule"	Yes	No	No	No
"Current State"	Yes	Yes	No	No
"Current & Upcoming"	Yes	Yes	Yes	No
"Current & Predicted"	Yes	Yes	No	Yes

The timing of monitoring is also important. Therefore, the extended monitoring will be tested at three different times within a mission: during a Pre-Flight action, an In-Flight action, and a Post-Flight action.

Figure 5-1 shows the timeline of the battery energy state, in kW-s, with both a nominal and degraded battery. Each vertical line indicates the time of a monitoring action. Prior to the flight, and before the vehicle expends any additional battery energy, the Pre-Flight monitoring action is paired with the Request Flight Authorization taskwork action and occurs once. After the flight, and before any possible charging may occur to increase the battery level, the Post-Flight monitoring action is paired with the Maintenance taskwork action and occurs once. During the flight, while the vehicle is actively expending battery energy, the In-Flight monitoring action is paired with the Flight Dynamics taskwork action. However, this taskwork action occurs with a frequency of 50 Hz, while the monitoring action occurs with a frequency of 1 Hz during the flight.

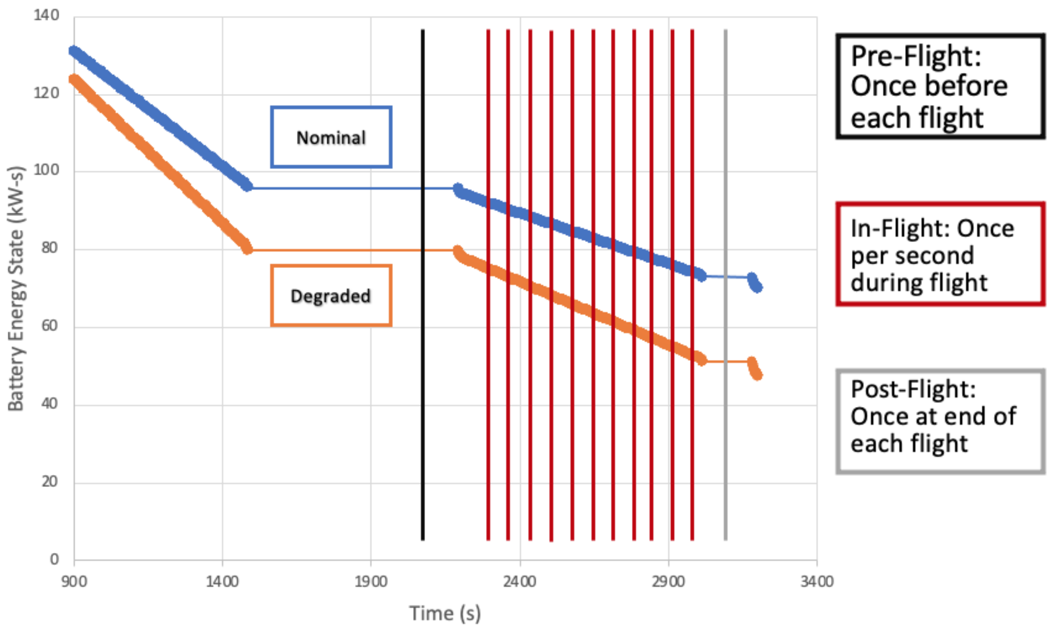

Figure **5-1**: Timeline of battery for UAS with nominal and degraded batteries with monitoring action timing.

Implementing these types of monitoring in WMC translates to a unique algorithm and set of resources that the monitoring agent will "get", as shown in Tables 5-2 through 5-4. Each of these resource "gets" may require information transfer between agents; this information transfer is considered teamwork.

Note that the "Current State & Upcoming Requirements" monitoring has the same resource requirements as the "Current State & Predicted" monitoring for the Post-Flight monitoring action. Based on its timing within the vehicle mission schedule, these two algorithms have the same impact. Therefore, for Post-Flight monitoring, the only information distributions that are investigated are "Current State" and "Current State & Predicted".

Table **5-2**: List of resources accessed and associated WMC Work Models for Pre-Flight Monitoring.

"Current State"	**"Current State & Upcoming Requirements"**	**"Current State & Predicted"**
Vehicle ID: UAS Mission ID: UAS Vehicle Number: Command/Control Battery Level: UAS Battery Capacity: UAS Home Flight Status: Mission	Vehicle ID: UAS Mission ID: UAS Vehicle Number: Command/Control Battery Level: UAS Battery Capacity: UAS Mission Status: Mission Mission Flight Status: Mission Return Flight Status: Mission Mission Required Energy: Mission Return Required Energy: Mission Home Flight Status: Mission	Vehicle ID: UAS Mission ID: UAS Vehicle Number: Command/Control Battery level: UAS Battery Capacity: UAS Mission Status: Mission Mission Flight Status: Mission Return Flight Status: Mission Vehicle Average Charge Rate: UAS Vehicle Relative Burn Rate: UAS Mission Projected Landing Energy: Mission Return Projected Landing Energy: Mission Mission Charge Time: Mission Return Charge Time: Mission

Table **5-3**: List of resources accessed and associated WMC Work Models for In-Flight Monitoring.

"Current State"	**"Current State & Upcoming Requirements"**	**"Current State & Predicted"**
Vehicle ID: UAS Mission ID: UAS Battery Level: UAS Battery Capacity: UAS	Vehicle ID: UAS Mission ID: UAS Battery Level: UAS Battery Capacity: UAS Mission Flight Status: Mission Return Flight Status: Mission Flight Phase: UAS Current Longitude: UAS Current Latitude: UAS Current Altitude: UAS	Vehicle ID: UAS Mission ID: UAS Battery Level: UAS Battery Capacity: UAS Mission Flight Status: Mission Return Flight Status: Mission Flight Phase: UAS Current Longitude: UAS Current Latitude: UAS Current Altitude: UAS Destination Longitude: Mission Destination Latitude: Mission Destination Altitude: Mission Origin Longitude: Mission Origin Latitude: Mission Origin Altitude: Mission Ascent Energy Burn: UAS Ascent Empty Energy Burn: UAS Descent Energy Burn: UAS Descent Empty Energy Burn: UAS Cruise Energy Burn: UAS Cruise Empty Energy Burn: UAS Mission Charge Time: Mission

Table **5-4**: List of resources accessed and associated WMC Work Models for Post-Flight Monitoring.

"Current State"	**"Current State & Upcoming Requirements"**	**"Current State & Predicted"**
Vehicle ID: UAS Mission ID: UAS Battery Level: UAS Battery Capacity: UAS Mission Flight Status: Mission Return Flight Status: Mission	Vehicle ID: UAS Mission ID: UAS Mission Assigned Vehicle: Mission Battery Level: UAS Battery Capacity: UAS Home ID: UAS Mission Flight Status: Mission Return Flight Status: Mission Mission Projected Landing Energy: Mission Return Projected Landing Energy: Mission	Vehicle ID: UAS Mission ID: UAS Mission Assigned Vehicle: Mission Battery Level: UAS Battery Capacity: UAS Home ID: UAS Mission Flight Status: Mission Return Flight Status: Mission Mission Projected Landing Energy: Mission Return Projected Landing Energy: Mission

Chapter 6

Results

This chapter summarizes the results of WMC simulations of the monitoring associated with different information distributions. As stated earlier, two simulations were conducted, one for a nominally performing vehicle battery and one for a degraded vehicle battery. Each examined whether monitoring would flag (correctly or incorrectly) a problem based on algorithms associated with different information distributions and monitoring occurring at different times within the operation.

6.1 Impact of Monitoring on Task load

Monitoring adds to the number of actions the agents need to execute, i.e., their task load. In WMC simulations, this can be seen in the resulting timeline of activity that each agent performs, referred to as the "action trace". As noted in chapter 4, the agents in these simulation runs do not have any limit on the task load they can simultaneously execute. This allows for a full picture of the total number of actions that monitoring adds to an operation.

In this case study's proposed allocation of authority and responsibility, three agents are assigned responsibility for actions, and therefore perform monitoring: the Command agent, the MissionOp, and the VehicleOp. The VertiportOp, while being assigned authority over many actions, is not allocated any responsibility and therefore does not complete any monitoring.

Figure 6-1 shows the action trace for the Command Agent. While the actions are spread consistently throughout the timeline for the day, the presence of monitoring adds a significant number of actions to the overall task load of the Command Agent. Higher lines in the action trace indicate more than one action occurring at a time. Some of these monitoring actions occur during gaps between taskwork actions, but other monitoring actions are expected to occur at the same time as taskwork, increasing the number of concurrent actions to a maximum of 7.

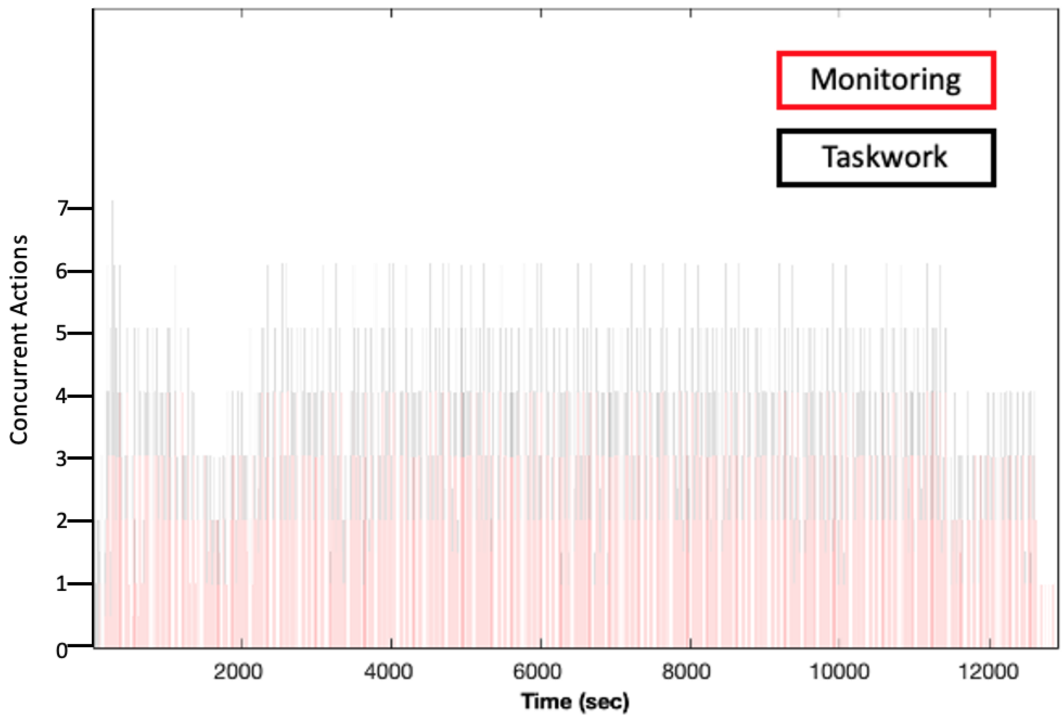

Figure **6-1**: Action trace for Command agent over one day's operations.

Figure 6-2 shows the action trace for the MissionOp agent, which is clearly dominated by monitoring actions. The MissionOp agent is the responsible agent for monitoring the Flight Dynamics taskwork and therefore conducts extended In-Flight monitoring with a frequency of 1 Hz during each vehicle flight. This creates a significant increase in overall task load for the agent. The spikes in monitoring actions also tend to coincide with spikes in taskwork actions.

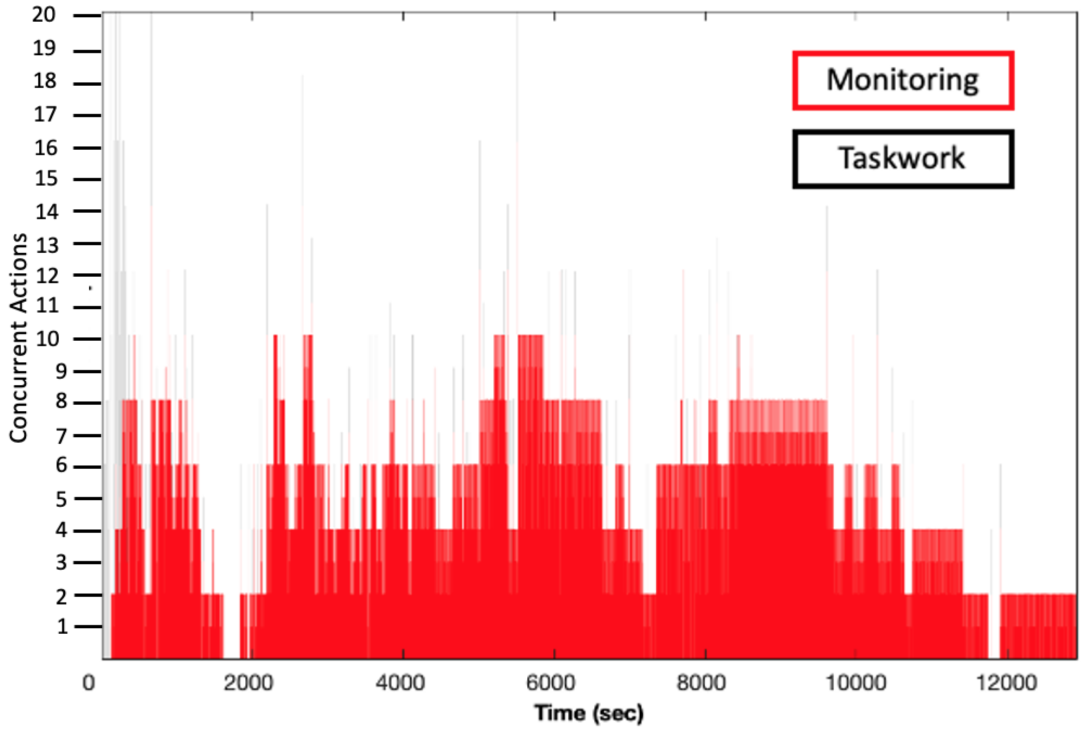

Figure **6-2**: Action trace for Mission agent over one day's operations.

Figure 6-3 shows the action trace for the VehicleOp agent. The VehicleOp agent is responsible for the Pre-Flight Extended monitoring action and is the assigned authoritative agent for the Flight Dynamics taskwork action, which occurs with a frequency of 50 Hz. This plot demonstrates the immense saturation of the VehicleOp task load due to the execution of the Flight Dynamics taskwork action. While there are some monitoring actions, taskwork very clearly dominates the task load for the VehicleOp agent.

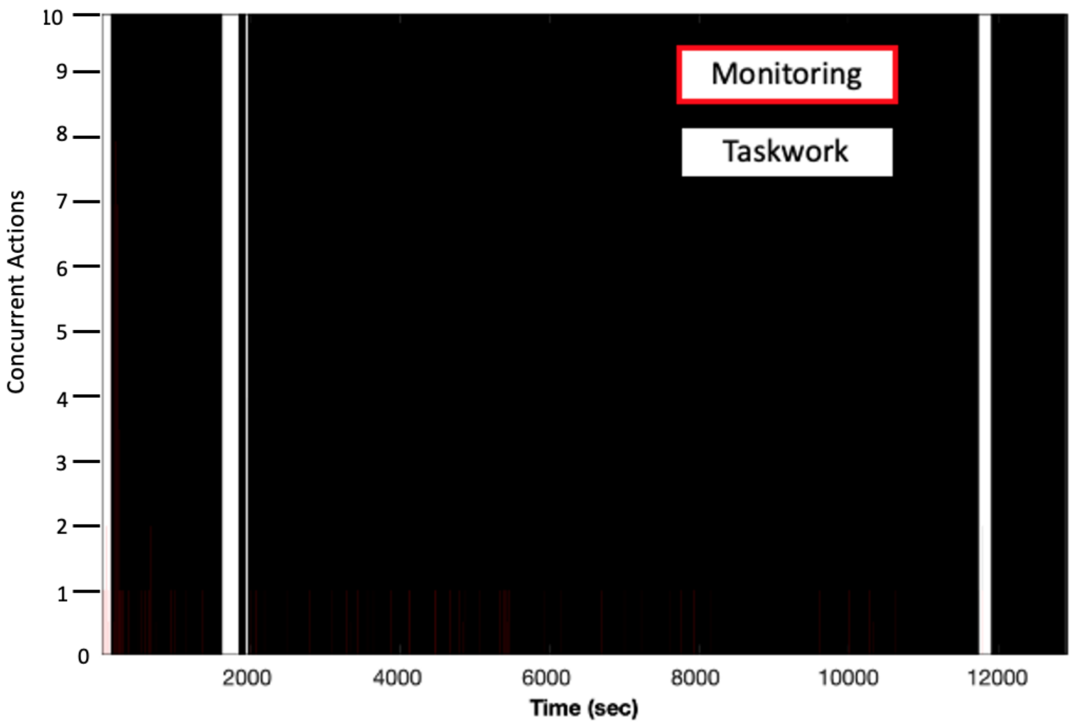

Figure **6-3**: Action trace for Vehicle Op agent over one day's operations.

Figure 6-4 summarizes the total number of actions for each of the agents that monitor. Using a logarithmic scale to accommodate the wide range of data, this table summarizes the monitoring and taskwork actions that each monitoring agent completes. The VehicleOp agent performs the most total actions, with those being almost entirely taskwork. The MissionOp agent performs the highest number of monitoring actions, due to its assignment as the responsible agent for the Flight Dynamics (In-Flight monitoring) action. The total number of taskwork and monitoring actions that an agent performs is directly influenced by the frequency at which the action is called within WMC. During a flight, the number of actions called also depends on the length of the flight. For example, the VehicleOp agent performs a very large number of taskwork actions due to the Flight Dynamics taskwork action being called with a frequency of 50 Hz. If the frequency of the action or the length of the flight are altered, the total number of taskwork actions performed by the VehicleOp agent changes as well. In the same manner, the MissionOp agent performs many monitoring actions due to its assignment as the monitoring agent for In-Flight monitoring, which occurs with a frequency of 1 Hz. If the frequency of the monitoring action or the length of the flight are altered, the total number of monitoring actions performed by the MissionOp agent changes.

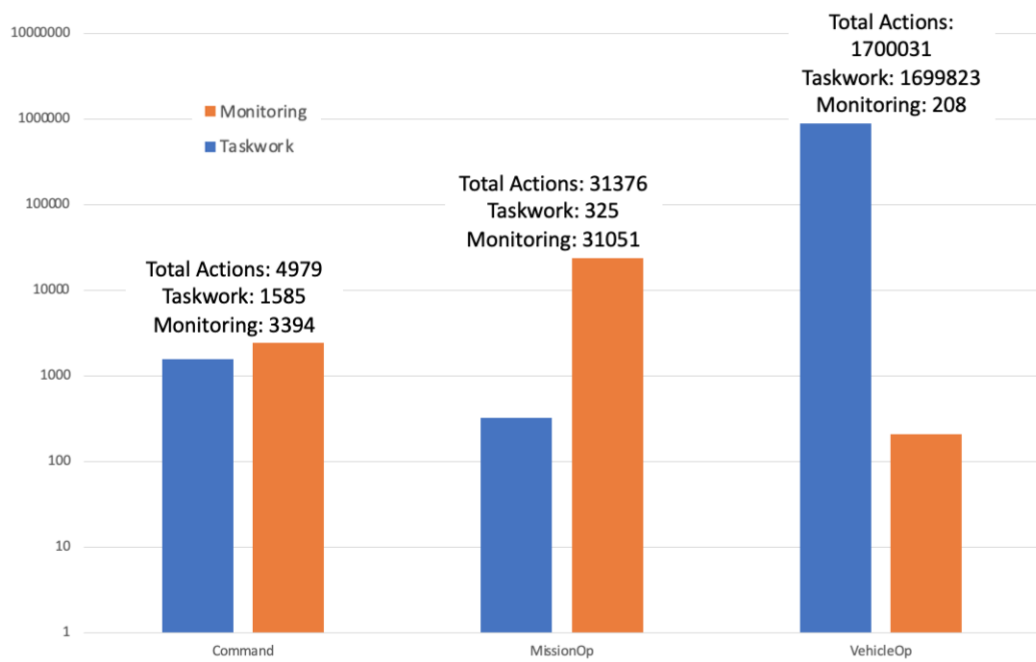

Figure **6-4**: Logarithmic plot of actions performed by each monitoring agent.

6.2 Impact of Monitoring on Information Transfer

An information transfer is defined as occurring when the monitoring agent “gets” a resource that has previously been “set” by another agent, implying that it needs to somehow be transferred between agents. Figure 6-5 illustrates the total number of information transfers that occur for each monitoring information distribution and the time of monitoring.

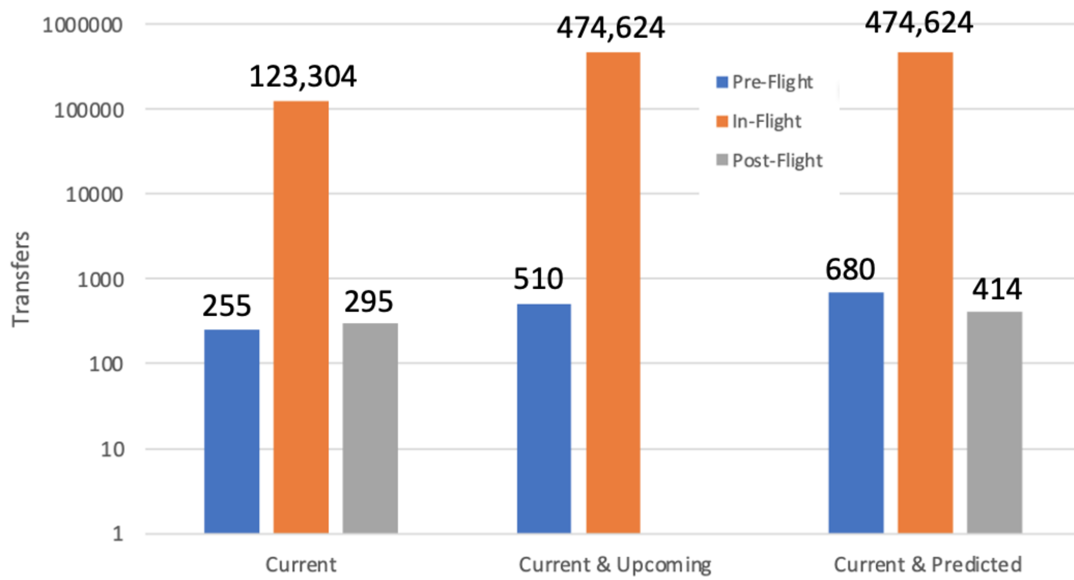

Figure **6-5**: Information transfers occurring for each information distribution.

The total number of information transfers in monitoring actions is influenced by two factors: information distribution and total number of monitoring actions called. For In-Flight monitoring in this case study, the total number of monitoring actions is influenced by the frequency of monitoring within the action, and the length of each flight within a mission. If the frequency of the monitoring actions or the flight lengths are altered, the total number of information transfers would be changed as well. To compare information transfers between different information distributions without considerations from total monitoring actions, table 6-1 lists the percentage increase in transfers that occur for information distributions beyond “Current State”.

Table **6-1**: Number of information transfers and percentage increase beyond the "Current State" information distribution.

	"Current State"	**"Current State & Upcoming Requirements" (% Increase from Current)**	**"Current State & Predicted" (% increase from Current)**
Pre-Flight: 60 flights	255	510 (100% increase)	680 (167% increase)
In-Flight: 60 flights at 1 Hz	123,304	474,624 (285% increase)	474,624 (285% increase)
Post-Flight: 60 flights	295	--	414 (40% increase)

Table 6-2 lists the total number of information transfers that are provided to each monitoring agent distributed by agent from which the monitoring agent is requesting resources. Resources transferred from the VehicleOp agent encompass the largest percentage of the total transfers, and resources transferred from the MissionOp and UAS default agents encompass the two smallest percentages of total transfers. Again, the number of information transfers is influenced by the information distribution and the total number of monitoring actions called.

Table **6-2**: Number of information transfers requested from each agent in support of monitoring.

	VehicleOp	**MissionOp**	**Command**	**VertiportOp**	**UAS Default**
"Current State"	92,655	0	31,140	0	59
"Current State & Upcoming Requirements"	400,738	310	30,996	12,204	60
"Current State & Predicted"	400,975	310	31,199	12,204	119

6.3 Impact of Information Distribution and Monitoring Time on Monitoring Performance

This case study applies two criteria to assess monitoring performance. The first criterion represents the situation when the vehicle does not have enough battery to complete the upcoming or current mission: This criterion is considered “TRUE” for any action time when the battery level is (or will be) lower than the mission’s required energy or return flight’s required energy. The second criterion represents detecting a degraded battery: This criterion is considered “TRUE” when the vehicle battery is degraded and is considered “FALSE” when the vehicle battery is operating nominally.

For each monitoring action algorithm, if the output exceeds an alerting threshold, it represents the monitoring agent detecting a problem. Therefore, a monitoring algorithm’s ratio of true negative and false positive detections (relative to a criterion that is FALSE), and of true positive and false negative detections (relative to a criterion that is TRUE), varies with its alerting threshold.

The performance of the monitoring across alerting thresholds is shown in System Operating Curves (SOC). The vertical axis of each SOC represents the observed frequency of a true positive detection by the monitoring agent given that the criterion is TRUE. The horizontal axis represents the observed frequency of a false positive detection by the monitoring agent given that the criterion is FALSE. Each run of the WMC simulation collects data for 5 vehicles collectively conducting 60 flights (a mission flight to the payload delivery location and a return flight to the home vertiport for each of the 30 missions). For every instance of a monitoring action, “TRUE” or “FALSE” alert values are determined for each data point according to each criterion, at varying alerting thresholds.

From these values, true and false positive and negative rates are determined. Using the rate values, X and Y data pairs were generated for each alarm threshold value within the data set, according to the following equations (NCSS):

$$Y = \frac{TP}{TP + FN}$$

$$X = 1 - [\frac{TN}{TN + FP}]$$

Each point along an SOC represents the monitoring algorithm performance observed at a different alerting threshold. High performance in monitoring a "TRUE" condition is represented by a high observed frequency of a true positive with a low observed frequency of a false positive.

The SOCs for Criterion 1, which offers the question, "Can we identify a vehicle without sufficient battery?", are shown in Figure 6-6 for all monitoring algorithms and times.

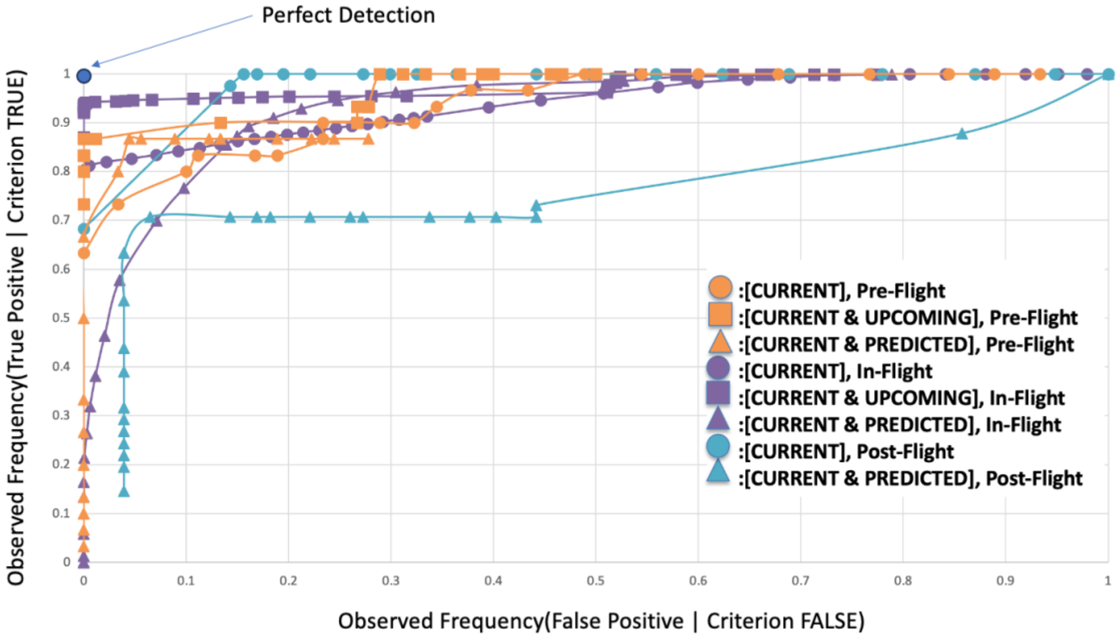

Figure **6-6**: SOCs for Criterion 1 for all monitoring algorithms and times.

None of the combinations of monitoring time and monitoring algorithm based on information distribution can achieve perfect detection performance. The monitoring algorithm that comes closest to a perfect detection uses the “Current State & Upcoming Requirements” information distribution, monitoring during the In-Flight action. This algorithm reaches a true positive rate of just under 94% before introducing false positives. The next closest monitoring algorithm uses the same information distribution but monitors during the Pre-Flight action, which reaches a true positive rate of just under 87% before introducing false positives.

Monitoring based on current state alone has a limit on the percentage of true positives achieved without lowering the detection threshold to the point that the algorithm also begins issuing false positives (i.e., its SOC deviates from the vertical axis). For monitoring based only on the “Current State” information distribution operating during the In-Flight action, the monitoring agent can achieve a true positive rate of approximately 81% before introducing false positives; using the same information distribution to monitor during the Pre-Flight action, the monitoring agent can achieve a true positive rate of approximately 63% before introducing false positives; using this information distribution to monitor during the Post-Flight action, the monitoring agent can achieve a true positive rate of approximately 68% before introducing false positives. Monitoring based on the “Current State” information distribution can achieve 100% true positives, however this would require lowering the alerting threshold to the point that monitoring will also trigger false positive alerts.

For this criterion, monitoring based on the “Current State & Predicted” information distribution has arguably the worst detection performance. Post-Flight monitoring with this information distribution always has a false positive rate of at least 4%, and demonstrates asymptotic behavior, lingering at a true positive rate of approximately 71% until the alerting threshold is lowered to the point of consistently generating false positives. Pre-Flight monitoring with this distribution has an upper limit on the number of achievable true positives of approximately 87%. In-Flight monitoring with this distribution begins introducing false positives very early, at a true positive detection rate of approximately 21%.

The SOCs for Criterion 2, which offers the question, “Can we identify a degraded vehicle battery?”, are shown in Figure 6-7. As discussed earlier, Criterion 2 is considered “TRUE” for the entire day with a degraded battery (even early in the day, before the degradation may be observable in its current state), and “FALSE” for the entire day with a nominal battery.

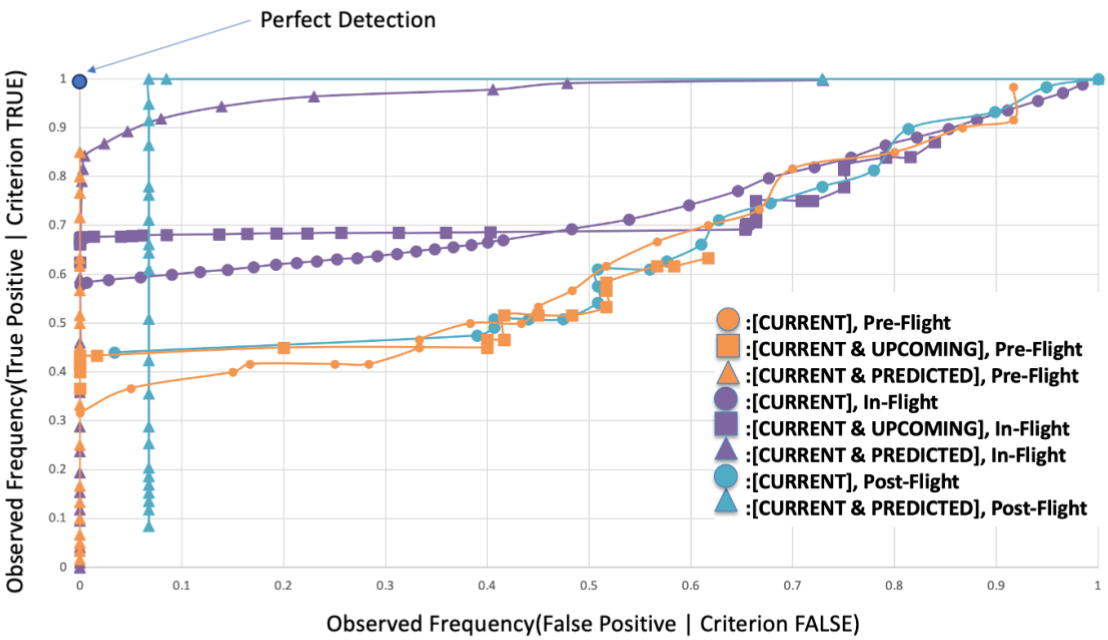

Figure **6-7**: SOCs for Criterion 2 for all monitoring algorithms and times.

In this case, monitoring using the “Current State & Predicted” information distribution has the best detection performance. This type of monitoring has the best performance when applied during the In-Flight or Pre-Flight actions, where it can achieve a rate of approximately 85% true positives without introducing false positives. When applied Post-Flight, this type of monitoring always has a false positive rate of at least 6.7%; however, this monitoring timing can achieve 100% true positive detections. In contrast, the monitoring algorithm based on other information distributions have true positive rates that are limited at reasonable alerting thresholds to between approximately 40% and 70%.

Figure 6-8 analyzes the SOCs of four In-Flight monitoring conditions in detail, focusing on the "upper left" portion of the graph, indicating high performance.

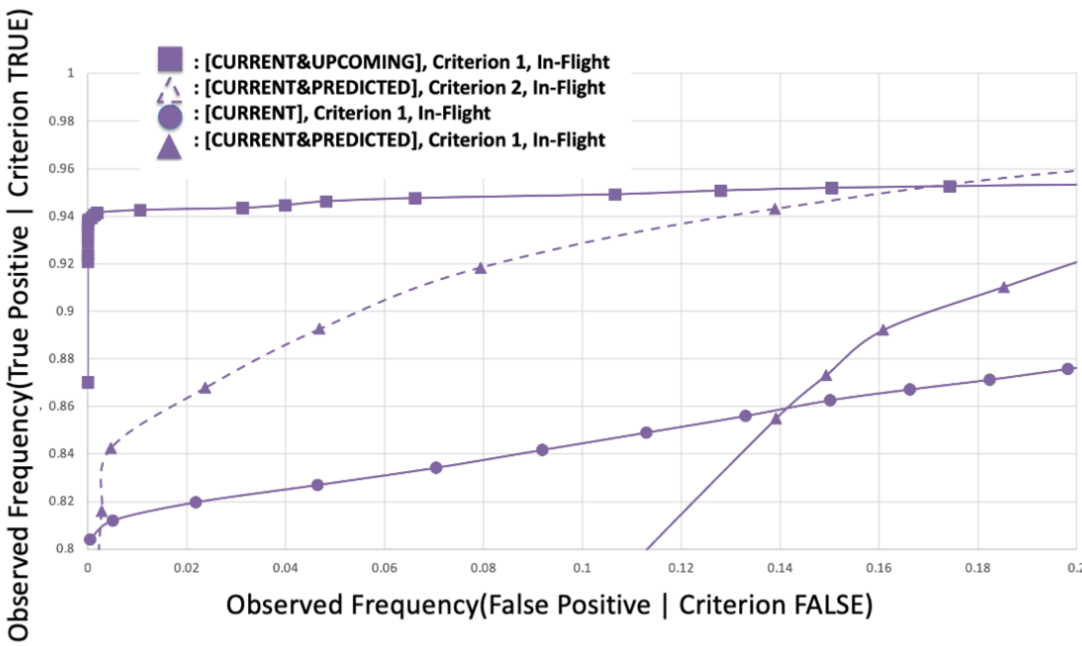

Figure **6-8**: Top left corner of SOCs for In-Flight monitoring.

The monitoring conditions that come closest to perfect true positive rates with minimal false positives depend on the criterion. When applying Criterion 1, which addresses whether the vehicle has sufficient battery energy for the upcoming flight, the monitoring algorithm using the "Current State & Upcoming Requirements" information distribution performs best. When applying Criterion 2, which addresses whether the battery is degraded, the monitoring algorithm using the "Current State & Predicted" information distribution performs best. Timing is also an important factor in monitoring performance: the best monitoring performance occurs during In-Flight monitoring.

Small changes in the alert threshold value can create significant changes in an SOC. Figure 6-9 illustrates the effect of different thresholds on the observed detection performance of the monitoring condition best able to detect Criterion 1.

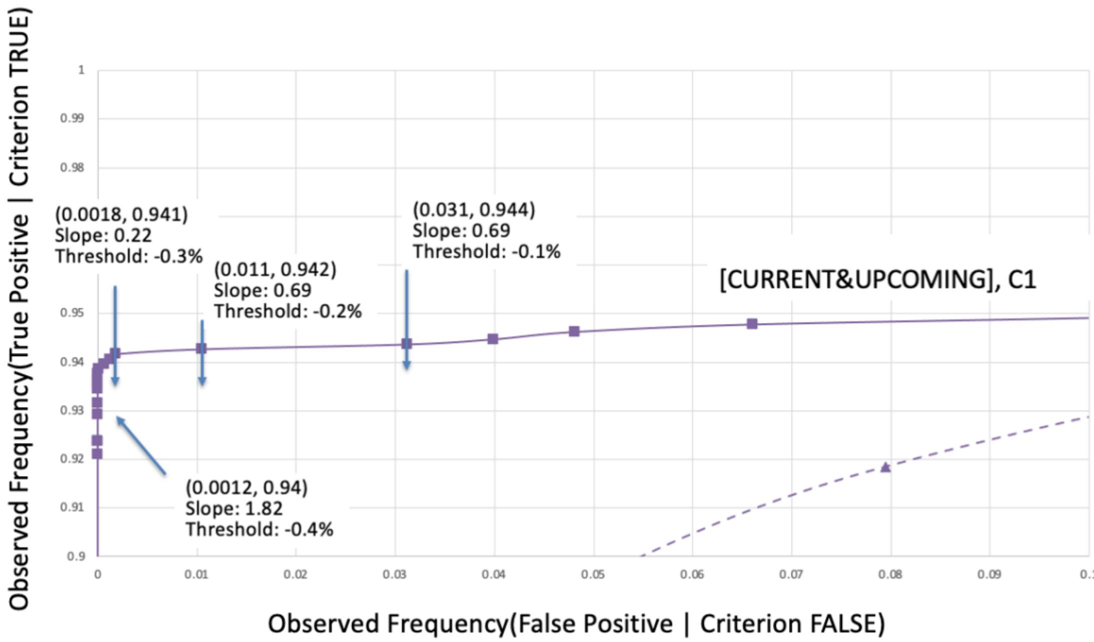

Figure **6-9**: Top corner of SOC for In-Flight monitoring based on "Current State & Upcoming Requirements" for Criterion 1 with different threshold points.

Within this SOC, the slope represents the inverse cost of a true positive increase, where the payment of that cost is in false positives. A high slope occurs when there is a low cost of true positives, or when true positives are increasing rapidly with very few false positives given as "payment". A low slope occurs when there is a high cost of true positives, or when many false positives must be "paid" to achieve an increase in true positives.

Chapter 7

Conclusion

Monitoring plays a crucial role in H-A teaming in aviation operations, fostering safety by identifying safety-critical situations. However, there has not been significant research regarding monitoring performance within distributed teams. Therefore, the objective of this thesis was to explore two significant factors of monitoring within a distributed team: when it is conducted, and the information distribution within the team that informs the monitoring agent. To explore these factors, this thesis applied an agent-based simulation, Work Models that Compute (WMC), to examine a Concept of Operations (ConOps) for Advanced Air Mobility (AAM). Within a day's operation involving several agents, 5 electric vehicles, and 30 missions totaling 60 flights, this case study introduced a degraded battery that would, if undetected, eventually result in the vehicle departing without sufficient energy to complete the flight. Agents involved in this case study were considered "perfect" and did not experience any task load limitations.

Monitoring was triggered in WMC by creating authority and responsibility mismatches in the action allocation. For each monitoring action, different information distributions were implemented in algorithms to demonstrate varying access to resources within a team.

Results highlight that, first, monitoring significantly increases the task load of the agents who perform it. For the Command and MissionOp agents, monitoring added a significant number of actions, with the MissionOp agent receiving the largest increase in task load of 31,051 monitoring actions added. Second, monitoring can significantly increase the amount of information that must be transferred between agents. The In-Flight monitoring action showed an increase of over 350,000 transfers when increasing the information distribution beyond the "Current State" information distribution, demonstrating an increase of 285%. As discussed in

Chapter 2, an allocation must be realizable with the associated teamwork (Feigh & Pritchett, 2014); information transfers are a form of teamwork, and therefore must be considered by system designers.

Third, examining the performance of the monitoring actions, ideal monitoring based on information with no bias or inaccuracy still presents limits on the rate of true positive and true negative detections that the monitoring algorithm can achieve. Certain combinations of when the monitoring is performed, and the information distribution provided to the monitoring agents, offer observed frequency performance rates that are little better than chance at many alerting thresholds. Even the best-performing combinations of monitoring timing and information distribution could not serve as perfect detectors, as they have to tradeoff between true positive and false positive detection rates and are often limited on how high of a true positive rate or how low of a false positive rate they can achieve. For each assessment criterion considered for monitoring analysis, the "best" performing monitoring actions were those that had access to information that was in line with the criterion. In addition to information distribution variation, the timing of monitoring affected performance for both alerting criteria. Monitoring action placement within a mission schedule influences the quality of resources available to the monitoring agent. Monitoring actions that are placed during times when current information is available will perform better than actions that do not have access to the most current information. However, it must be considered that while In-Flight monitoring may have performed better than other monitoring times according to the SOCs, if a problem is detected during a flight, the flight must be aborted, potentially causing additional safety concerns. Operationally, the ideal time for a monitoring agent to detect a degraded battery is before a flight takes off.

The monitoring performance observed in this thesis was determined from one simulated day's operations. Specific effects, such as the asymptotic limits on achievable true positive and false positive rates, were potentially influenced by the nuances of the day's schedule. For example, monitoring based on current battery level for the criterion that the battery is too low to finish a mission will tend to detect cases more before longer flights than shorter flights. Future simulations that consider different patterns of operations, such as a different length or number of flights, may observe different rates of true and false positives.

Analyzing monitoring performance in the context of alerting thresholds provides additional insights regarding future design considerations. When constructing missions containing monitoring, system designers must consider the effect of alert thresholds on a monitoring agent's ability to identify safety-critical scenarios. The threshold at which automated tools will alert monitoring agents to safety-critical circumstances will greatly impact the agent's performance. Having too high a threshold may hinder detection and therefore lead to missed events. However, having too low a threshold may result in a higher occurrence of false positives, which also have their own cost and safety concerns.

This work identified crucial elements of mission design that must be considered when implementing monitoring in ConOps. This monitoring may be done by human agents or automated agents. Either way, this thesis demonstrated how the information available to the monitoring agent, and the time at which monitoring is conducted, can affect monitoring performance, as well as how the addition of monitoring will affect overall task load and information transfer requirements. Further, this thesis examined only one of many types of hazards that would need to be monitored for within a distributed H-A team; additional hazards must be considered in ConOps design.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779. https://doi.org/10.1016/0005-1098(83)90046-8

Bass, E. J., & Pritchett, A. R. (2008). Human-automated judge learning: A methodology for examining human interaction with information analysis automation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 38*(4), 759-776. https://doi.org/10.1109/TSMCA.2008.923068

Bellenkes, A. H. (1999). *The use of expert pilot performance models to facilitate cockpit visual scan training.* Champaign, IL: Doctoral dissertation, University of Illinois Urbana Champaign. https://hdl.handle.net/2142/82277

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68*(7), 569-579. PMID: 9215461.

Billinghurst, S. S., Morgan, C., Rorie, R. C., Kiken, A., Bacon, L. P., Vu, K.-P. L., Strybel, T. Z., & Battiste, V. (2011). Should students learn general air traffic management skills before NextGen tools. *Proceedings of the human factors and ergonomics society annual meeting, 55*(1) (pp. 128-132). Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/1071181311551027

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bolstad, C. A., Endsley, M. R., Howell, C. D., & Costello, A. M. (2002). General aviation pilot training for situation awareness: An evaluation. *Proceedings of the human factors and ergonomics society annual meeting, 46*(1) (pp. 21-25). Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/154193120204600105

Bolstad, C., Endsley, M. R., Costello, A. M., & Howell, C. D. (2010). Evaluation of compter-based situation awareness training for general aviation pilots. *The International Journal of Aviation Psychology, 20*(3), 269-294. https://doi.org/10.1080/10508414.2010.487013

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Individual and group decision making: Current issues* (pp. 221-246). Lawrence Erlbaum Associates. ISBN: 9781138876248

Carbonell, J. R. (1966). A queueing model for many-instrument visual sampling. *IEEE Transactions on Human Factors in Electronics, 7*(4), 157-164. https://doi.org/10.1109/THFE.1966.232984.

CAST. (2014). *Airplane state awareness joint safety analysis team: Final report analysis and results.* Commercial Aviation Safety Team. https://www.cast-safety.org/pdf/JSAT-ASA_FinalReport_June2014.pdf

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research, Vol. 2* (pp. 1-12). Leeds: Emerald Group Publishing. ISBN: 0-7623-0864-8

Dearden, A., Harrison, M., & Wright, P. (2000). Allocation of function: Scenarios, context and the economics of effort. *International Journal of Human-Computer Studies, 52*(2), 289-318. https://doi.org/10.1006/ijhc.1999.0290

Deloitte Consulting LLP. (2020). *UAM vision concept of operations (ConOps) UAM maturity level (UML) 4 Version 1.0.* NASA Technical Reports Server. https://ntrs.nasa.gov/api/citations/20205011091/downloads/UAM%20Vision%20Concept%20of%20Operations%20UML-4%20v1.0.pdf

Entin, E. E., & Entin, E. B. (2001). Measures for evaluation of team processes and performance in experiments and exercises. In *Proceedings of the 6th international command and* *control research and technology symposium* (pp. 1-14). Annapolis, MD. http://dodccrp.org/events/6th_ICCRTS/Tracks/Papers/Track6/090_tr6.pdf

Federal Aviation Administration. (1996). *The interfaces between flightcrews and modern flight deck systems.* Washington, D.C.: FAA. https://www.tc.faa.gov/its/worldpac/techrpt/hffaces.pdf

Feigh, K. M. (2011). Incorporating multiple patterns of activity into the design of cognitive work support systems. *Journal of Cognition Technology and Work, Volume 13*(4), 259-279. https://doi.org/10.1007/s10111-010-0165-0

Feigh, K. M., & Pritchett, A. R. (2014). Requirements for effective function allocation: A critical review. *Journal of Cognitive Engineering and Decision Making, 8*(1), 23-32. https://doi-org.ezaccess.libraries.psu.edu/10.1177/1555343413490945

Ferrari, F., Spillmann, K. P., Knecht, C., Bektas, K., & Muhlethaler, C. (2018). Improved pilot training using head and eye tracking system. *Eye tracking for spatial research, Proceedings of the 3rd international workshop* (pp. 58-63). Zurich: ETH Zurich. https://doi.org/10.3929/ethz-b-000222482

Flin, R., Martin, L., Goeters, K.-M., Hormann, H.-J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (Non-Technical Skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety, 3*(2), 95-117. https://psycnet.apa.org/record/2003-08985-001

Froger, G., Blättler, C., Dubois, E., Camachon, C., & Bonnardel, N. (2018). Time-interval emphasis in an aeronautical dual-task context: A countermeasure to task absorption. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 60*(7), 936-946. https://doi.org/10.1177/0018720818783946

Hauland, G. (2008). Measuring individual and team situational awareness during planning tasks in training of en route air traffic control. *The International Journal of Aviation Psychology, 18*(3), 290-304. https://doi.org/10.1080/10508410802168333

Hollenbeck, J. R., Ilgen, D. R., Sego, D. J., Hedlund, J., Major, D. A., & Philips, J. (1995). Multilevel theory of team decision making: Decision performance in teams incorporating distributed expertise. *Journal of Applied Psychology*, *80*(2), 292-316. https://doi.org/10.1037/0021-9010.80.2.292

Hollnagel, E., & Woods, D. D. (2006). *Joint cognitive systems: Foundations of cognitive systems engineering.* Boca Raton, FL: Taylor & Francis Group. ISBN-10: 0849328217

Holt, R. W., Boehm-Davis, D. A., & Hansberger, J. T. (2001). *Evaluation of proceduralized CRM at a regional and major carrier.* Fairfax, VA: George Mason University.

Hormann, H. J., Banbury, S. P., Dudfield, H. J., Lodge, M., & Soll, H. (2004). Effects of situation awareness training on flight crew performance. In S. Banbury, & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 213-232). Ashgate. ISBN: 0754641988

Hutchins, E. (1996). *Cognition in the wild.* Cambridge, MA: The MIT Press. ISBN: 9780262581462

Hutchins, E. J., Holland, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human-Computer Interaction, 1*(4), 311-338. https://doi.org/10.1207/s15327051hci0104_2

IJtsma, M. (2019). *Computational simulation of adaptations of work strategies in human-robot teams.* Atlanta, GA: Doctoral dissertation, Georgia Institute of Technology. http://hdl.handle.net/1853/61783

IJtsma, M., Ma, L. M., Pritchett, A. R., & Feigh, K. M. (2019). Computational methodolgy for the allocation of work and interaction in human-robot teams. *Journal of Cognitive*

Engineering and Decision Making, 13(4), 221-241. https://doi-org.ezaccess.libraries.psu.edu/10.1177/1555343419869484

Kaber, D. B., Omal, E., & Endsley, M. R. (1999). Level of Automation Effects on Telerobot Performance and Human Operator Situation Awareness and Subjective Workload. In M. W. Scerbo, & M. Mouloua (Eds.), *Automation technology and human performance: Current research and trends* (pp. 165-170). Mahwah, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-3135-5.

Kearns, S. (2011). Online single-pilot resource management: Assessing the feasibility of computer-based safety training. *The International Journal of Aviation Psychology, 21*(2), 175-190. DOI: 10.1080/10508414.2011.556499

Knecht, C. P., Muehlethaler, C. M., & Elfering, A. (2016). Nontechnical skills training in air traffic management including computer-based simulation methods: From scientific analyses to prototype training. *Aviation Psychology and Applied Human Factors, 6*(2), 91-100. https://doi.org/10.1027/2192-0923/a000103

Lee, J. D. (2006). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1570-1596). Indianapolis: Wiley. DOI:10.1002/0470048204

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 35*(10), 1243-1270. https://doi.org/10.1080/00140139208967392

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., Linton, P. M., Wierwille, W. W., Zaklad, A. L., Bittner, Jr., A. C., & Wherry, R. J. (1989). *Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies.* Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences. https://apps.dtic.mil/sti/pdfs/ADA212879.pdf

Ma, L. M., IJtsma, M., Feigh, K. M., Paladugu, A., & Pritchett, A. R. (2018). Modelling and evaluating failures in human-robot teaming using simulation. *IEEE aerospace conference*. DOI: 10.1109/AERO.2018.8396581

Malakis, S., Kontogiannis, T., & Psaros, P. (2014). Monitoring and evaluating failure-sensitive strategies in air traffic control simulator training. *Proceedings of the 7th international conference on PErvasive Technologies Related to Assistive Environments - PETRA* (pp. 1-6). New York: Association for Computing Machinery. https://doi.org/10.1145/2674396.2674462

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human Factors, 49*(1), 57-75. https://doi.org/10.1518/001872007779598037

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors, 38*(2), 311-322. https://doi.org/10.1177/001872089606380211

Mumaw, R. J., Billman, D., & Feary, M. S. (2020). *Analysis of pilot monitoring skills and a review of training effectiveness.* NASA Technical Reports Server. https://human-factors.arc.nasa.gov/publications/NASA_TM_20210000047.pdf

Mumaw, R. J., Haworth, L. A., & Feary, M. S. (2019). *The role of alerting system failures in loss of control accidents.* Hampton, VA: NASA STI Program. https://ntrs.nasa.gov/api/citations/20190004927/downloads/20190004927.pdf

NCSS. One ROC curve and cutoff analysis. In NCSS, *NCSS statistical software* (pp. 546.1-546.34). NCSS. https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/One_ROC_Curve_and_Cutoff_Analysis.pdf

NRC, (2003). *Securing the future of U.S. air transportation: A system in peril.* Washington, D.C.: The National Academies Press. https://doi.org/10.17226/10815

NTSB (1994). *Safety study: A review of flightcrew-involved, major accidents of U.S. air carriers, 1978-1990.* Washington, D.C.: NTSB. https://www.ntsb.gov/safety/safety-studies/Documents/SS9401.pdf

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, Vol. 39*(2), 230-253. https://doi.org/10.1518/001872097778543886

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

PARC/CAST. (2013). *Operational use of flight path management systems.* Performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team flight deck automation working group. https://www.faa.gov/sites/faa.gov/files/aircraft/air_cert/design_approvals/human_factors/OUFPMS_Report.pdf

Potter, B. A. (2011). *Effects of online training on aircrew monitoring behaviors: A field study*. Daytona Beach, FL: Master's thesis, Embry-Riddle Aeronautical University. https://commons.erau.edu/edt/117

Pritchett, A. R. (2001). Reviewing the Role of Cockpit Alerting Systems. *Human Factors in Aerospace Safety, 1*, 5-38. https://doi.org/10.4324/9781315259482-11

Pritchett, A. R., Feigh, K. M., Kim, S. Y., & Kannan, S. K. (2014). Work models that compute to describe multiagent concepts of operation: Part 1. *Journal of Aerospace Information Systems, 11*(10), 610-622. https://doi.org/10.2514/1.1010146

Pritchett, A., Portman, M., & Nolan, T. (2018). *Research & technology development for human autonomy teaming final report: Literature review and findings from stakeholder interviews.* Hampton, VA: NASA Langley Research Center.

Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors, 43*(4), 641-674. https://doi.org/10.1518/001872001775870386

Sarter, N. B., & Woods, D. D. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1926-1943). Wiley. ISBN-13: 978-0471116905

Sperling, B. K., & Pritchett, A. R. (2011). Complementary information distribution to improve team performance in military helicopter operations: An experimental study. *International Journal of Aviation Psychology, 21* (4), 375-396. https://doi.org/10.1080/10508414.2011.606756

Villa, A. (2022). *Computational modeling of monitoring within human-autonomy teams with novel delegations of authority and responsibility.* State College, PA: Master's thesis, Pennsylvania State University . https://etda.libraries.psu.edu/files/final_submissions/26690

Vu, K. P., Kiken, A., Chiappe, D., Strybel, T. Z., & Battiste, V. (2013). Application of part-whole training methods to evaluate when to introduce NextGen air traffic management tools to students. *The American Journal of Psychology, 126*(4), 433-447. https://doi.org/10.5406/amerjpsyc.126.4.0433

Wiener, E. (1985). Beyond the sterile cockpit. *Human Factors, 27*(1), 75-90. https://doi.org/10.1177/001872088502700107

Wiener, E. L., & Curry, R. E. (1980). Flight deck automation: Promises and problems. *Ergonomics, 23*(10), 995-1011. https://doi.org/10.1080/00140138008924809

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors, 45*(3), 360-380. https://doi.org/10.1518/hfes.45.3.360.27250

Woods, D. D. (1985). Cognitive technologies: The design of joint human-machine cognitive systems. *AI Magazine, 6*(4), 86-92. https://doi.org/10.1609/aimag.v6i4.511

Zacharias, G. L. (2015). *Advancing the science and acceptance of autonomy for future defense systems.* Testimony to House Armed Services Committee. https://docs.house.gov/meetings/AS/AS26/20151119/104186/HHRG-114-AS26-Wstate-ZachariasG-20151119.pdf