Investigating the Tradespace between Increased Automation and Optimal Manning on Aircraft Carrier Decks

by

Weston Ross

Department of Mechanical Engineering and Materials Science
Duke University

Date: ______________________

Approved:

____
Mary Cummings, Supervisor

____
Michael Zavlanos

____
George Konidaris

Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Mechanical Engineering and Materials Science
in the Graduate School of Duke University
2016

Abstract

Investigating the Tradespace between Increased Automation and Optimal Manning on Aircraft Carrier Decks

by

Weston Ross

Department of Mechanical Engineering and Materials Science
Duke University

Date: ______________________
Approved:

____
Mary Cummings, Supervisor

____
Michael Zavlanos

____
George Konidaris

An abstract of a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Mechanical Engineering and Materials Science
in the Graduate School of Duke University
2016

Abstract

With increasing prevalence and capabilities of autonomous systems as part of complex heterogeneous manned-unmanned environments (HMUEs), an important consideration is the impact of the introduction of automation on the optimal assignment of human personnel. The US Navy has implemented optimal staffing techniques before in the 1990's and 2000's with a "minimal staffing" approach. The results were poor, leading to the degradation of Naval preparedness. Clearly, another approach to determining optimal staffing is necessary. To this end, the goal of this research is to develop human performance models for use in determining optimal manning of HMUEs. The human performance models are developed using an agent-based simulation of the aircraft carrier flight deck, a representative safety-critical HMUE. The Personnel Multi-Agent Safety and Control Simulation (PMASCS) simulates and analyzes the effects of introducing generalized maintenance crew skill sets and accelerated failure repair times on the overall performance and safety of the carrier flight deck. A behavioral model of four operator types (ordnance officers, chocks and chains, fueling officers, plane captains, and maintenance operators) is presented here along with an aircraft failure model. The main focus of this work is on the maintenance operators and aircraft failure modeling, since they have a direct impact on total launch time, a primary metric for carrier deck performance. With PMASCS I explore the effects of two variables on total launch time of 22 aircraft: 1) skill level of maintenance operators and 2) aircraft failure repair times while on the catapult (referred to as Phase 4 repair times). It is found that neither introducing a generic skill set to maintenance crews nor introducing a technology to accelerate Phase 4 aircraft repair times improves the average total launch time of 22 aircraft. An optimal manning level of 3 maintenance crews is found under all conditions, the point at which any additional maintenance crews does not reduce the total launch time. An additional discussion is included about how these results change if the operations are relieved of the bottleneck of installing the holdback bar at launch time.

Contents

List of Tables

List of Figures

Acknowledgements

There are many people to who I owe gratitude for supporting me during my graduate career to date and during the writing of this thesis.

Firstly, I'd like to thank my girlfriend Katy. Without her consistent support, encouragement, and sympathetic ear, none of this would have been possible.

Secondly, I'd like to thank my friends for their guidance and assistance in reviewing my thesis.

Thirdly, I'd like to thank my post doc, Alex, for his guidance in developing this research. His advisement was crucial to completing this thesis.

Lastly, I'd like to thank my family, who have been supportive of me throughout these many years in college and graduate school. Without y'all I couldn't have made it nearly a far as I have.

Thank you all.

1

Introduction

Partially and fully autonomous agents are rapidly being adopted in high-risk or high-performance environments for their capabilities in performing tasks quickly, consistently, and to remove humans from dangerous environments. For example, astronaut-robot teams are being developed in which the human operator oversees various types of autonomous robotic agents during a lunar construction task requiring the movement of large panels, welding, and weld inspection [13]. Alongside manned aircraft and human operators, autonomous aircraft are landing and taking off from Naval aircraft carrier decks [19], a historic first. Mining operations, due to their high risk nature, benefit in both safety and efficiency from incorporating autonomous agents and human operators into the same teams, however this has not been without difficulty [30].

Inadequate training, slow-human acceptance of automated systems, and over-reliance on automated systems are among the most common problems in heterogeneous manned-unmanned environments. As autonomy increases, the optimal role of the human in the system will likely change, and determining the appropriate role for the human must consider the impact to the system for both safety and efficiency. Human performance modeling (HPM) and agent based modeling (ABM) [1], also known as agent based simulation (ABS) [2], are common approaches to modeling complex environments [4] including heterogeneous manned-unmanned environments (HMUE) in which humans and autonomous agents work together or in close proximity.

1.1 Motivation

It is necessary to consider human performance factors in HMUEs because humans are a necessary component of complex systems due to their flexibility and experience, but they are also error prone especially when fatigued. An astronaut may be more likely to make a welding mistake causing injury or costly repair time if he is fatigued from hours of moving metal sheets and welding tasks. Working in human-robot teams, autonomous agents can perform low level tasks such as moving heavy materials and welding, allowing the astronaut to save energy and oversee the high-level task planning operations, something to which a human is much better suited than robots [13]. This is not to say autonomy is always preferable to manpower, or vice versa; there is a tradeoff, where robots offer strength and reliability, humans offer intelligent reasoning and flexibility. While the field of artificial intelligence is continually improving algorithms that reason under uncertainty for robots, the frame problem [17] will always be a hurdle due to rare, unexpected events or contingency situations. The frame problem is the problem of needing to define all possible actions and consequences under every condition, something even the most experienced carrier deck operator would not be capable of doing. Therefore, complex and or unpredictable environments exist that are not ready or may never be ready to fully remove the human operator.

Particularly important areas for human performance and manning considerations are HMUEs in which humans work closely with potentially dangerous systems, which may be partially or fully autonomous. Human performance modeling (HPM) in simulated HMUEs can be used to determine the optimal role of the human operators and autonomous agents in that environment given a performance metric and objective function. In many environments there is pressure to reduce costs while maintaining or improving performance and safety. The Navy's Operations and Support (O&S) costs are about 50% of the total operating budget [56]. So, a reduction in manpower or more efficient usage of manpower is of high priority and offers the potential for a large reduction in cost. Furthermore, having fewer people working in hostile environments, such as the carrier deck, reduces the number of safety incidents. This does not necessarily mean that 'minimum staffing' is desirable because it can lead to under-preparedness of the system, particularly in times of high workload. Minimum staffing refers to having the fewest number of human operators required to maintain the function of a system or environment at a particular performance level. However, staffing and performance requirements may change drastically from nominal to off-nominal conditions, potentially creating deficits in performance under minimum staffing approaches. For example, the Navy has found that such practices can lead to an "overall decline in shipreadiness" [41, 45], meaning that in the event of a contingency operation or surge in operation requirements as occurs in wartime, the crewmen aboard would not be sufficient manpower to run the ship.

1.2 Research Approach

In an ABM, each agent's behaviors and interactions are described individually at an agent level and then simulated in an environment with many agents to study the system-level dynamics that emerge [1, 21]. This allows for the investigation of how changes in behavior at the agent level will affect the overall performance of the system. The aircraft carrier deck environment is highly heterogeneous, having numerous types of aircraft, operators, and equipment on deck. For these reasons, I have chosen the US Navy aircraft carrier deck as a representative HMUE for testing human performance models. Through ABM, each of the carrier deck entities, i.e. planes, catapults, operators, and equipment, can be treated as independent “agents”, where each agent has certain properties or tasks associated with it. Agent based modeling allows for each type of agent’s behaviors to be define independently of the other agent types and to have stochasticity in their actions, both of which are necessary to modeling aircraft carrier deck operations.

In order to test my human performance models of operators on the aircraft carrier deck, I am developing a simulation called the Personnel Multi-Agent Safety and Control Simulation (PMASCS). PMASCS is an agent-based simulation of the aircraft carrier deck launch operations and is written in the object oriented language of Java. These performance models are designed to be modular, so that they may be used in other simulation environments as well. PMASCS is an extension of an existing carrier deck operations simulation called Multi-Agent Safety and Control Simulations (MASCS) [46, 47]. MASCS was developed to compare the effectiveness of several unmanned vehicle control architectures on carrier deck operations. It focused primarily on the movement of the aircraft on the flight deck and the single operator type that guides this movement, the aircraft director. PMASCS extends MASCS to include a detailed model of scheduled and unscheduled aircraft maintenance as well as the human performance models for the seven major operator types on deck:

1. aircraft directors and catapult officers

2. ordinance officers

3. equipment officers

4. fueling officers

5. plane captains

6. safety officers, and

7. maintenance and catapult crews,

as well as a detailed model of scheduled and unscheduled aircraft maintenance. Chapter 3 describes the changes made to MASCS in extending it to PMASCS.

1.2.1 Model Development and Validation

Verification and validation is a necessary and crucial part of developing any simulation. The validation of PMASCS and the related human performance models has occurred primarily through observation of activities on the carrier deck, subject matter expert (SME) review, and building from the calibration and validation from the original MASCS simulation. The general strategy employed for this effort was to calibrate and validate these models on an agent-by-agent basis using total completions times as a metric, then at a system level using total sortie launch time as a metric. For the carrier deck, agents behaviors are built on probabilistic distributions derived from SME input as well as Naval carrier deck operations documents, statistics, videos, and personal observations. I have visited two aircraft carrier decks to interview operators and observe operations. I first visited the *USS Theodore Roosevelt (CVN 71)* while she was docked in Norfolk, VA and received a tour of the vessel and interviewed commander Ronald Rancourt. Then, I visited at sea the *USS John C. Stennis (CVN 74)* to observe flight operations in action and interview various crewmen. It was during these visits that I obtained the observations and SME input to develop my model. During development, the model was continually validated and adjusted based on SME feedback.

1.3 Preliminary Exam Research Questions

It is the goal of this Master's thesis to answer the following questions and support or reject the following hypothesis for the development of the human performance models and with PMASCS.

1. Which parameters (e.g. task times and distributions, fatigue, and error probabilities) are sufficient to model human operator performance on a carrier deck, such that the models are acceptable to SMEs and are able to be validated by sortie rate statistics?

2. How do manning levels of aircraft maintenance crewmen affect the performance of aircraft carrier deck sortie rates given increased automation and skill level?

   (a) Hypothesis 1a: Current operations of aircraft carriers are conservatively staffed, and the nominal total launch time can be achieved with fewer maintenance crews than are in the standard complement.

   (b) Hypothesis 1b: Broadening the skill level of maintenance crews to allow them to maintain both E-2 and F-18 aircraft will improve the total launch time compared to current operations for very small manning levels.

   (c) Hypothesis 1c: Increased automation to reduce the Phase 4 failures maintenance time will significantly improve the launch rates, since aircraft with Phase 4 failures can cause big queues behind the catapult they are failed on.

3. What is the minimum number of maintenance operators required on deck to maintain the nominal total launch time, given a standard complement of other personnel?

(a) Hypothesis 2a: Both generic skill levels and increased automation will allow for fewer maintenance crews on deck than the standard complement while maintaining the nominal total launch time.

(b) Hypothesis 2b: Generic skill set will have more of an influence on the total launch time than the increased automation in Phase 4 failure maintenance, so the fewest number of maintenance crews will be achievable with a generic skill set.

1.4 Outline of Master's Thesis

This Master's thesis contains the following chapters.

- Chapter 2: Literature Review — This chapter is intended to give the reader a introduction to the bodies of work and modeling methods related to this thesis work. Specifically, it gives an introduction to:
  - Optimal Manning / Optimal Manning for Naval Operations
  - Carrier Deck Operations
  - Human Performance Modeling
  - Agent Based Modeling
  - MASCS
- Chapter 3: MASCS to PMASCS — This chapter describes the development of PMASCS from MASCS, including the human performance and aircraft maintenance models.
- Chapter 4: Simulations, Results, and Analysis — This chapter presents the four experiments designed for testing the hypothesis and provides a statistical analysis of results for each using one way ANOVAs to show statistical significance

across manning levels. The final section, Section 4.5, provides a comparison of results across experiments and manning levels

- Chapter 5: Conclusions and Future Work — This chapter describes the recommendations for improvements to staffing of maintenance crewmen on the carrier deck. It also discusses potential future work as well as possible extensions to other domains.

- Appendix A — This appendix includes tables for the results of the pairwise comparisons across manning levels presented in Chapter 4.

- Appendix B — This appendix includes a description of the Kruskal Wallis H test and assumptions along with distributions of total launch times for each of the four experiments.

- Appendix C: Automated Holdback Bar Pareto Analysis — This appendix provides an alternative Pareto frontier analysis to Section 4.5 when the holdback bar installation is hastened through automation.

1.5 Summary

This work focuses on the development of human performance modeling in heterogeneous manned-unmanned environments. Specifically, HPMs of maintenance operators and aircraft failure models are presented that I have designed. These models are then used to extend an existing agent based simulation of the aircraft carrier deck to conduct various analyses of the interactions between agents and the models' affect on system performance. The simulation of the aircraft carrier deck environment is modular and could be adapted to work in many HMUEs and future areas of research.

2

Literature Review

This chapter is intended to provide the reader with an introduction and background on the techniques that are used in this research project. The overarching topic is human performance modeling and optimal manning, with a focus on carrier deck operations. It begins with introduction to optimal manning in general, then focuses in on optimal manning efforts already made by the US Navy. Section 2.2 describes carrier deck operations during a launch cycle and the role of each of the operator types. Following that, Section 2.3 contains introductions to the many disciplines within human performance modeling, identifying which types of models are specifically found in this work. The models developed in this work are implemented in an agent based model called PMASCS, an extension of an agent based model called MASCS, so I conclude with an introduction to agent based modeling and MASCS in Section 2.3. Chapter 3 elaborates on the differences and additions made to MASCS in extending it to be PMASCS.

2.1 Optimal Manning

Human capital, in many industries and organizations, is the most valuable resource but oftentimes comprise the highest cost. Particularly in service based and manufacturing industries, the number of employees on hand directly affects the efficiency or quality of service provided. That said, there is frequently a diminishing return on additional staff members when over-staffed levels are reached, at which point the quality of service does not necessarily improve with an additional staff member, though it may cost more. So, it is not surprising that finding the optimal balance between costs and performance in terms of the number of staff is of interest in many industries. The answer, though, to the optimal staffing level cannot be disjoint from the question of "In what way is it optimal?" Optimal manning is also referred to throughout the literature by any of the following: work force scheduling, personnel scheduling, optimal staffing, and minimal staffing. This thesis uses the term 'optimal manning.' Similar to an optimal controls problem, optimal manning requires a cost function and a variable or many variables that should be optimized upon.

Optimal manning is already being applied to many fields. To increase productivity and reduce labor cost, optimal manning was tested in the operating room [18]. Studies of manning efforts of nurses on patient healthcare in hospitals [39] determines which types of nurses, when given more hours, have the greatest benefit to the patients. This study is incomplete, though, in the sense that it only considered one metric for optimization, the patient health. Oftentimes it is necessary to consider other metrics such as profit margins, budgetary constraints, regulations, or maximum human workload when determining an optimal manning solution. Aimed at determining the optimal number of registered nurses, licensed practical nurses, and nursing aids, [18] does just this. Their metrics are cost of ulcer care treatment and prevalence and severity of the ulcers in patients, arguably two competing variables.

Some studies have found that there is little to no analytical support for environments that have implemented optimal manning. For example, in Australia, helicopter emergency medical services removed the physician from the helicopter in order to optimally man the helicopter emergency medical service (HEMS), a controversial move that has no literature support [44].

Optimal Manning in the Navy

In 1996 the Navy implemented for the first time its Smart Ship technology on the USS Yorktown. Smart Ship was an effort to increase automation onboard Navy vessels with the goal of reducing the crew complement and improving quality of life for those on the ship. The introduction of Smart Ship allowed for a $2.8 million per year savings by reducing the complement by 15 percent [56]. Human error in dividing by zero while using the Smart Ship technology caused a cascading failure of systems on board, leaving the ship dead in the water for some time. Just a few years later, the Navy commissioned a study of effectiveness of this technology [56] that ultimately determined that the "ship design process [should] include Human Engineering so that optimal human/system performance is achieved with as few Sailors as possible." This statement embodies the next decade of the Navy's approach to optimizing their ship complements through what are essentially minimal-manning techniques. The next generation of ships, the DDG 1000 class, utilized human factors to drive the design of the ship, considering the capabilities and limitations of the ships crewmen. The design relies heavily on automated systems and has a greatly reduced crew complements of less than half that of similar ships [42]. This is just one example of human factors engineering greatly impacting the future of Naval operations. Much of the design principals were developed from observations in usability testing (UT) of shipboard systems. These tests measured metrics of usability including user reaction times, error rates, number of actions, task completion times, and cognitive load.

Many of these metrics are common amongst the models presented within this work as well as other human factors models, which are presented in the Section 2.3.

One important consideration that is not addressed in this modeling work is the readiness of the ship to handle contingency operations and large-scale failures such as plane wrecks or fires. Hiltz found that, for the Canadian navy, the complement size can be significantly reduced without adversely affecting the crews ability to handle contingency operations [20].

Criticisms of lack of ship readiness that followed in the decade after implementing Smart Ship has lead to this work as another means of studying the human factor on aircraft carriers, a step towards developing an improved method of optimal ship manning. In a letter to congress in 2011 [41], S. Pickup, Director of Defense Capabilities and Management, comments on a report the Navy submitted in 2010 with a study of the effectiveness and impact of their optimal manning efforts which began in 2000. The study found that "the decreased manning aboard the Navy's surface combatant ships due to the optimal manning initiative, which removes sailors from ships through efficiency initiatives, contributed to declines in material readiness and an unmanageable workload burden on crews" [41]. The report collected information on many ship statistics, including the total manning numbers and skill training for maintenance crewmen. This thesis studies the impact of both of these factors on the PMASCS simulation of carrier deck operations, as well as how a technology improvement will affect the total launch times.

2.2 Carrier Deck Operations and Personnel

Aircraft carriers are the largest warships of a naval fleet and allow for aircraft to be deployed practically anywhere in the world without requiring a land-based landing strip. The United States Navy maintains 10 aircraft carriers, each 333 meters long and nuclear powered. These are the largest carriers in the world and are referred to as super carriers, each capable of carrying up to 72 aircraft, consisting of both fixed wing and rotary wing aircraft. This work focuses on the launch cycle of fixed wing aircraft, specifically the F-18 fighter jet and the E-2 Hawkeye. From sunrise to sunset, operators are working on deck to provide support for the launching and maintenance of aircraft. A launch cycle is completed about every two hours. A launch cycle may also be referred to as a sortie, and the terms are used interchangeably in this thesis. A sortie consists of a launching and recovery cycle of a group of aircraft for a mission. The group size can vary from just a few aircraft up to 35 aircraft for the largest launch, which may occur during wartime operations. The following sections describe the three main phases of a sortie: Pre-Launch Preparation, Catapult Launch, and Recovery.

2.2.1 Pre-Launch Preparation

Prior to launch, all aircraft must be prepared for their launch cycle. This includes getting fueled, attachment of ordnances, performing any pre-startup maintenance, and fixing any failures that occur at startup. Much of this can be done in parallel, as depicted in Figure 2.1 and in Figure 2.2. The Operator Activity Timeline, shown in Figure 2.2, depicts the most active times of each of the operators during the launch cycle. An elevated level of activity means that they have a well-defined task responsibility during that time period. The vertical blue and green lines indicate the beginning of the PMASCS simulation and the launching of the first aircraft, respectively. Beyond that, operators are busied with prepping and launching the remaining aircraft in the launch cycle. This is not depicted in the figure. Note the maintenance crew activity timeline. They are most active during the first hour of the launch cycle and periodically active in the second half. The periodic activity in the second half is because of unforeseen aircraft failures requiring the maintenance crew. This is discussed in more detail in Chapter 3 and Chapter 4. Once the aircraft is prepped and the chocks are removed by the Chocks and Chains crewmen (which are a subset of Equipment Operators), the Plane Captain allows the aircraft to taxi to launch. The cyclic arrows shown between the Chocks and Chains and the aircraft in Figure 2.1 indicate that the plane may be chocked multiple times prior to launching due to unforeseen failures or delays. Aircraft Directors guide the aircraft to its assigned catapult once it is ready for launch.

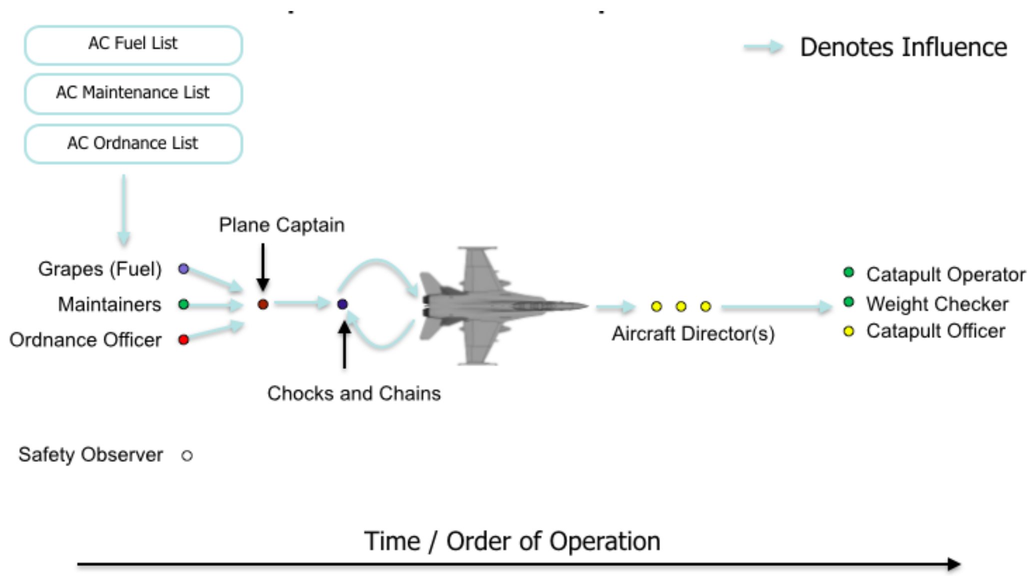

FIGURE 2.1: Launching Preparation Timeline

2.2.2 Catapult Launch

There are four catapults that may concurrently be used for launching of aircraft. Figure 2.3 shows the locations of the catapults, and Figure 2.4 shows two aircraft being prepped for launch on catapults 1 and 2. Catapults 1 and 2 cannot be simultaneously launched, but one aircraft may be prepped while the other is in use. The same is true for catapults 3 and 4. The only time catapults 2 and 3 cannot be simultaneously launched is if either one is launching an E-2. The E-2 wingspan is too large to all for both to be concurrently launched. Catapults 3 and 4 are in the landing strip of the carrier, so they cannot be used if aircraft are being recovered.

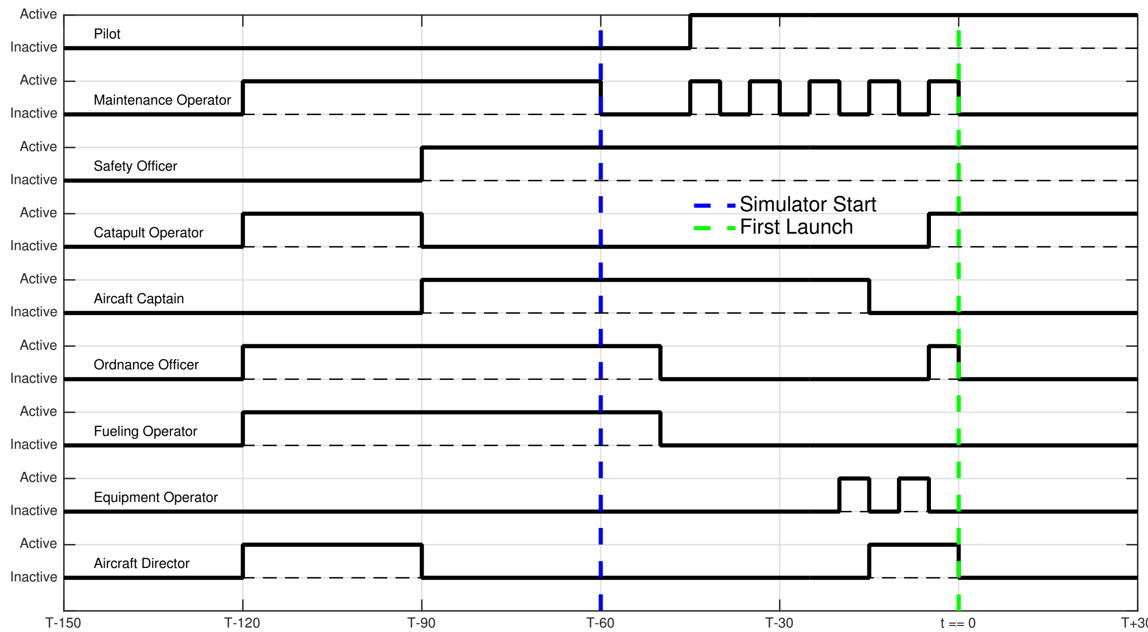

FIGURE 2.2: Operator Activity Timeline

This work focuses on aircraft launch cycles without recoveries with all four catapults in operation. Once on the catapult, the weightboard checks and attachment of the holdback bar are performed by the catapult crews. Weightboard checks are a confirmation between the catapult crew and the pilot on the estimated weight of the aircraft. This weight is used to calibrate the pressure of the steam-powered catapults. Underpowered catapults can result in a failed launch and a lost aircraft. Overpowered catapults can result in an unnecessarily high acceleration stressing both the aircraft and the pilot. Future aircraft carriers are replacing the steam powered catapult with an electromagnetic aircraft launch system (EMALS) to improve reliability and control [9].

2.2.3 Recovery

Aircraft are recovered by approaching from the aft of the carrier deck and landing on the strip containing catapults 3 and 4. There are three or four cables that span the width of the landing strip which are used to arrest the aircraft. On the back

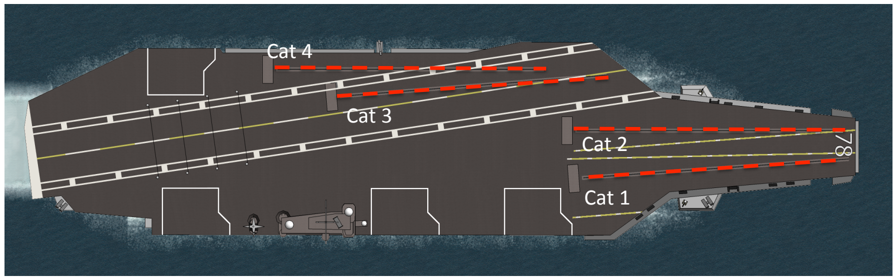

FIGURE 2.3: Catapult Locations

FIGURE 2.4: Photo of Catapults 1 and 2 from Tower on CVN-74

of each aircraft is a hook that is lowered for landing. The pilot must touch down in such a way as to drag the hook along the deck surface so it may catch a cable, preferably cable 3. It is not infrequent for them to miss the cables all together and take off again. It also occurs occasionally that the cable may brake, requiring that the aircraft launch again or else roll off the front of the carrier into the ocean. For this reason, the pilots hit full throttle immediately upon touching down and maintain the full throttle until instructed by a landing officer that the aircraft is appropriately arrested and can do reduce the throttle. The aircraft is then disconnected from the arresting cable and taxied to a parking spot in preparation for the next sortie. An aircraft at the end of its landing arrestment on the *USS Stennis* is shown in Figure 2.5.

FIGURE 2.5: Recovery of Aircraft on CNV-74

2.2.4 Operators

There are many operators and operator types involved in launch operations on the carrier deck. Many of them are depicted in Figure 2.1 and can be seen in Figure 2.6. At the helm of the launch operations is the air officer, more commonly known as the "air boss." The air boss is responsible for all launch operations and all other operators on deck. They ensure that maintenance, launch, and recovery plans are executed according to that day's flight plan. They communicate with the other operator crews over radio from the tower where they track aircraft locations on a 1:192 scaled version of the carrier deck called a Ouija board. Outside on the deck itself, there are two categories of operators: squadron and deck. The squadron operators are organized into squadrons of aircraft and are responsible for all the maintenance and preparation of the aircraft. Deck crews are not aircraft specific, and are tasked with taxiing aircraft across the deck and running the catapult and recovery operations. In Section 3.2.3, the squadron operators are discussed in detail as they, rather than the deck operators, are the focus of this research project. Table 3.3 lists each of the squadron operator types and their roles in the launch cycle.

Figure 2.6: Squadron Crews during Pre-Launch Preparation

Three important deck operators are the aircraft directors, landing signal officers, and arresting gear officers. The aircraft directors are responsible for directing taxiing aircraft. During launch operations, this is from the aircraft's parking spot to their assigned catapult, at which the launch crews take point ownership of the aircraft. Landing signal officers (LSO) assist in the recovery of aircraft by monitoring the state of an approaching aircraft and communicating with the pilot using radios and a visual indicator called the "meatball," more technically named the Improved Fresnel Lens Optical Landing System (IFLOLS) [54]. Approaching pilots use the meatball to determine their glide angle with respect to the deck to assist them in landing. A LSO visually examines the approach and calls off the landing if they think it is not satisfactory for a safe recovery. Once recovered, the arresting gear officers (AGO) reset the cables and prepare the recovery system for the next landing. Though these operators are crucial for carrier deck landing operations, since this thesis focuses only on the launch cycle, they are not considered in this work. Only human performance models of squadron crews are presented.

2.3 Human Performance Modeling

Central to human factors research are human performance models (HPM). Human performance modeling is used to describe the behaviors of a person in a particular environment or while performing a particular task. Its goal is to understand how a person's behavior can impact a system, or how a system can impact a person, and to thereby drive system design. For example, in designing an interface for control and monitoring of a nuclear power plant, it would be useful to know how the placement of buttons and screens could affect the ability of the operator to detect problems and fix them. Using a HPM in conjunction with a model of the power plant control station could help identify design trade-offs between, say, displaying the most information and being able to highlight a dangerous failure when it occurs without it being hidden in clutter.

Human performance modeling must cover a vast field of topics, because human behaviors are naturally complex. Grossly speaking, the majority of HPM's can be binned into seven major areas of research: Models of Perception and Attention, Models of Visual Search, Workload Modeling, Action or Motor Control Modeling, Cognition and Decision Making, Emotion Modeling, and Anthropometric and Biomechanical [6]. Each of these is discussed briefly in the remainder of this chapter.

2.3.1 Models of Perception and Attention

In many environments, performance and safety is a function of the perception and attention of the human operator. Failure to perceive a warning light indicating a failure on an aircraft could result in a fatal crash. A train engineer losing focus while controlling a train can, and has, resulted in disaster [10]. So, it is comforting to know that perception and attention models have been studied at length and are well established. Signal detection theory (SDT) [31], a model of signal detection, is the most commonly used modeling framework in human factors.

SDT models whether or not a person will detect a signal of interest in a noisy background. Will the operator see the flashing light on the cluttered display of the aircraft cockpit or not? Given that the signal is present, if the operator correctly identifies it, the event is called a hit; otherwise the event is called a miss. Given that the signal is not present, if the operator determines that there is a signal this is called a false alarm; if he determines there is no signal, this is called a correct rejection. These concepts are generally based in statistical decision theory. Classical Neyman-Pearson statistics [49, 52] distinguish between Type I and Type II errors. Type I errors are referred to as false alarms and Type II errors are detection misses. Signal detection theory has been used in many fields to determine operator attentiveness air traffic control [3], hazard detection of drivers [61], and air-to-air combat fighting simulators [11]. Common amongst these studies is the comparison of attentiveness between novice and expert operators, finding that expert operators are much more attentive in general but may be more risky.

Models of Visual Search

Models of visual search are another commonly used modeling framework in human factors. Visual search models answer questions such as "Where will the human operator look next?"; "How fast can the pilot locate this dial on the flight control screen?"; and "How long will the pilot have to look at this dial to read the value?" Models that answer these types of questions can drastically change the design of a user interface if they suggest the usability of the interface is visually difficult. It could suggest, for example, that some signals should be more easily identifying than other signals on a cluttered display if they are more critical, or that the display should be made less cluttered. Much of this is already present in modern day interfaces, but as technology progresses, systems become more complex and the number of systems a single operator is required to monitor increases, and much more information is available than a person can process. A person tasked with guidance and navigation of a swarm of drone aircraft could not possibly monitor all of the state variables of each aircraft simultaneously, so the interface must bring to attention the most important ones in a way that the person can receive that information accurately and quickly. When the question of "How much information can someone process?" arises, we begin to move into the realm of workload modeling.

Workload Modeling

Workload modeling is based on the idea that there is a limit to the capacity of a human to perform certain tasks, either cognitively [40] or physically. It is easy to determine how much weight one could lift once, say, on the bench press at the gym. It's a more difficult to determine how many times someone should lift a ten pound box every day at a distribution center without injuring themselves in the long term. Workload becomes less well defined when you ask questions such as "What is the cognitive capacity of a this person?" Even though human factors as whole has not come to a clear definition of what constitutes cognitive workload and by what metrics it should be measured, much human factors research is currently being done that is producing useful findings. Air traffic controllers operate at the limits of their cognitive workload [33] because of the constant requirement of attention to their duties. Conversely, train engineers can either be physically tired, mentally tired, or bored, leading to a lack of appropriate response and a high error rate [15].

2.3.2 Action or Motor Control Modeling

Whereas models of visual search pertain to information reaching the human operator, models of action and motor control pertain to the reaction of the operator to that information. For example, the difference from the time an operator receives a visual cue, such as a flashing light, to the time they respond with an action, such as a button actuation, is called their response time. This can be approximated with an action model using the Hick-Hyman law [53]. The Hick-Hyman law estimates the response time as a function of the information entropy of the interface, generally based on the time it takes to resolve a binary decision tree. This law, however, assumes that all tasks are equally difficult given that a decision to complete the task has been made. This is not always the case. A smaller button that is further away would be more difficult to press quickly than a large button that is near to the hand. Fitts' Law [53] models this by determining an index of difficulty of a pointing or pressing action as a function of the distance to the target and the width of the target. In effect, it estimates how much positional uncertainty a human operator has to resolve to perform an action.

Most actions of interest, though, are not simply pressing buttons or pointing at stationary targets, they are much more complex. One of the most commonly modeled human operator actions is the steering of a car, but of course other, more complex actions exist such as pole-vaulting and rock climbing. The task of modeling these types of actions essentially becomes a feedback control problem in which the closed loop transfer function of not just a mechanical system, but also the transfer function of a human-machine system must be determined. Both the system and the human have their own internal dynamics that much be captured and which may be coupled. It has been shown that about 200ms is a threshold for human reaction times to visual stimuli [59], and therefore the minimum lag in the system due to the human response, given that there is no anticipatory information available to the human. One such model is the optimal control model [27], which can be likened to model predictive control or receding horizon control [34]. In each of these control schemes, a system dynamical model is contained in the controller (or human operator). The dynamical model is used to estimate the future state of the system given a set of future control inputs, and once a desirable trajectory is obtained only one time step of the optimal control sequence is applied and the trajectory is recalculated based on the current state. It can be argued that since human operators have intuition of the dynamics of a system based on experience, they too have an internalized dynamical model of the system that can be used to estimate response and determine a desirable control input. This is the basis of the optimal control model. Once developed, these types of models may be used to estimate human workload and operator attentiveness [5]. Some more advanced techniques of machine learning using artificial neural networks have been used to build motor control models of professional drivers [62]. As intelligence and autonomy becomes more prevalent, the role of the operator is shifting more and more into a supervisory role, requiring models of cognitive tasks.

2.3.3 Cognition and Decision Making

There is an entire section of human factors engineering that concerns itself with cognitive modeling and decision making models, and broadly speaking you could refer to this as a subset of artificial intelligence. When artificial intelligence is being discussed, naturally the Turing test comes to mind. The Turing test is a test to determine whether or not a machine is exemplifying a certain level of intelligence or cognition, and, ultimately, whether it is at the intellectual level of a normal human or better. If a machine passes the Turing test, then it is possible someone would not be able to tell the difference between a human's responses or the machine's responses. In this section I do not discuss whether or not a particular machine or algorithm is capable of passing the Turing test, but rather I present two of the fundamental cognitive models that may form the basis of more complex models that could one day stand up to the Turing test. These two models are the GOMS model and the lens model. In a sense, the models' aims are simply to mimic the ability of cognition in a particular action or decision-making process, not to simulate human intelligence in its entirety.

The Goals, Operator, Methods, and Selection rules (GOMS) model is used for modeling actions that someone already knows how to do [7, 25]. The actions themselves may consist of a set of smaller actions that need to be completed in sequence. A simple version of the GOMS model is the keystroke-level model (KLM) [7]. A KLM is a set of actions with associated completion times that must be completed in order to fulfill some larger task. For example, a train conductor's individual action list may consist of the following: a) Turn on/off the train bell (800ms); b) Look for approaching vehicles at road crossing (3 seconds); c) signal the horn (2 seconds). The task to be completed could consist of the following set of actions a-b-c-b-c-a, totaling 11.6 seconds. This could be a (albeit simple) KLM model of crossing a road for a train engineer. The operator modeling presented in this work could be considered a stochastic KLM, where each task's completion time is drawn from a distribution rather than a constant value. GOMS assumes that tasks must be completed in a serial fashion and are error-free. Extensions of GOMS such as Cognitive Perception and Motor GOMS (CPM-GOMS) [24] address this and admit parallel streams of actions which do not interfere, such as observing the oncoming traffic while actuating the horn.

Many actions require some form of real time decision making and are not necessarily routine, a core assumption of GOMS modeling. So, many models have been developed to determine which decision an operator will make. Three such models are *subjective prospect theory*, *expected utility theory*, and *elimination by aspects model.*

These models all broadly fall under the field of decision theory. *Subjective prospect theory* effectively models the expected utility of an action in consideration of non-deterministic actions and outcomes. At the center of this approach is the model's ability to capture the utility function of the person appropriately and the (perceived) likelihood of outcomes following from actions, which can be likened to a state transition matrix for Markov decision processes. When the probabilities are known by the operator, a theoretically optimal decision can be determined and this falls under *subjective prospect theory*. *Subjective prospect theory* is often used to determine the optimal decision strategy for use in comparing it against a human's decision strategy, allowing for a determination of how off-optimal their decisions may be. This is appropriate when the outcomes are easily quantifiable, but quite difficult when they are not. Consider someone choosing the next best move in a card game. The optimal decision is whichever would give them the greatest chance of winning considering all future possible actions of theirs and their opponents. Now consider someone at a department store choosing which shirt to by. What is the optimal decision? Aside from price, quantitative comparison is difficult to justify, so qualitative reasoning might be used. Based on color, size, fabric pattern, and fit, the consumer may rank the importance of each quality and begin reducing the number of candidate shirts by elimination. Models that predict this type of decision making are called *elimination by aspects (EBA) models* [60]. EBA models have been used to model consumer choices of coffee ground brands as a function of promotions [12]; consumer choices of freight train services [63]; and risky choices such as those found in gambling in games with double-dice roles [43]. Often times the emotional state of someone can have a significant influence on their decision making process. For this, human factors teams with psychology in developing emotional models of humans.

2.3.4 Emotion Modeling

As social beings, we are naturally good at picking up cues that indicate someone's emotional state. We have learned to identify subtleties in someone's body language, facial expressions, as well as not so subtle cues in one's decision making or expression of emotion through word choices. It's something we are inherently good at as people. And we care about identifying someone else's emotional state because it can help explain their decisions and actions. The decision making process can be heavily influenced by one's emotional state [22] and even cause the decisions to be irrational. So, it is important understand the emotional state of an operator, and the advantages to a system in being able to do so are enormous. This is why we develop emotional models. What is most difficult about this is achieving a measurement or ground truth for one's emotional state on which to build the models because it is not directly measureable. One field of research is aimed at doing so via facial expressions [16, 55]. In these works, image processing is used to identify the emotional state of children under various conditions and this data is then used to help determine if the child is likely to develop autism. Such models would enable computer systems to connect more fully to their human operators and the human-machine interface would become much more connected. Eliciting an emotional response from a user or changing the interface for the user under certain emotional states could dramatically improve (or at least avoid degradation of) the system as a whole. Ferrari SpA, for example, is researching emotional detection to incorporate into their vehicles. If the driver is found to be in an unsafe emotional or mental state, such as too tired, the vehicle reduces the power availability and steering responsiveness to all for a safer ride.

2.3.5 Anthropometric and Biomechanical

Anthropometry is the science of measurement of the physical properties of a person, e.g. the length of their foot or their center of mass [23]. Such measurements can be used to build models for ergonomic design of human-machine interfaces, clothing design, and furniture design, for example. Biomechanics is the study of their movement and underlying mechanisms that allow for such movement. [35] presents a model for heart rate responses to exercise intensities. Models such as this, when coupled with workload models, could provide the means to determining the optimal amount of weight someone should be moving at their job. Furthermore, heart rate is directly measurable, so the system can be real time and feedback controlled to maximize the output of the human-machine system, such as a cyclist.

2.4 Agent Based Modeling

Agent based modeling (ABM), also known as agent based simulation (ABS) [4, 21] or agent based modeling and simulation (AMBS), is a powerful discrete-event modeling technique used to study the dynamics of complex dynamical systems. In such systems, the dynamics can be described as behaviors of individual "agents" or components of the system. Traffic flow on a freeway, for example, is a system in which the overall system behavior is dependent on the individual dynamics of each vehicle and is a problem particularly well suited for ABM [4, 37]. In essence, the dynamics of an individual agent and their dynamics with respect to other agents and their environment are coded into each agent. The agents are then linked together in such a way that they can interact within the environment, and then the system is simulated. The user can observe the individual interactions and also the emergent behaviors of the entire system [4]. Agent based modeling is particularly useful for studying the dynamics of systems that would otherwise be too difficult to analyze analytically, too expensive to validate *en vivo*, or too rare to validate empirically. Simple ABM's will have fixed definitions of behaviors. Advanced ABM's can include techniques to model agent learning and adaptation through the use of neural networks and evolutionary algorithms [14, 32].

There are four main components necessary to any ABM. These are outlined in [1]. The first is the model or observer. The model is what builds and controls each of the components of the system, oftentimes a video game engine such as Golden T Game Engine (GTGE) [57]. Secondly, an environment in which the agents exist and interact must be built. This is also known as a space, field, or world. This is the environment in which agent's location relative to each other or their boundaries can be defined. In the case of an aircraft carrier deck, these are the edges of the carrier and particular regions, such as the catapult, that may change an agent's behaviors. Thirdly, each of the agent's behaviors must be defined. These behaviors are also called strappables, actions, or procedures. This is the heart of the simulation and special care must be taken in defining these behaviors because they will ultimately be what contributes to the emergent behaviors of the system, that which is usually the focus of the simulation to begin with. Lastly, a queue of agents and actions, also known as a schedule or a forever procedure, is kept and iterated over to move the simulation forward. The queue could be part of the game engine and consist of a list of agents. The queue is iterated through, calling upon each agent to update its state accordingly during that time step based on their specific action queue. In this fashion, time is stepped forward and the model is propagated. Agents may be responsible for tracking their time history of state variables themselves, or a separate component of the simulation may be built for this specific purpose. An optional component, if visualizations are desirable for the modeling, is a graphics controller, which may be part of the game engine. Additionally, for usability and debugging, a display of environmental information as well as agent's states may be included.

ABS has been used in many areas to study the dynamics of populations and systems such as group and social dynamics [51], the stock market [26], supply chains [58], and aircraft carrier decks [46, 47]. In [26], Johnson discusses the Santa Fe Artificial Stock Market (ASM) [29], an ABM developed to model the financial market with heterogeneously trading “agents” that can learn and modify their trading behaviors. It has been used to study the dynamics of proposed trader behavior models and validate them against empirical trends.

In [46, 47], Ryan et al. describe an ABM called Multi-Agent Safety and Control Simulation (MASCS). PMASCS, presented in this thesis, is an extension of MASCS. MASCS is discussed in detail in Section 2.5.

2.5 MASCS

The Multi-Agent Safety and Control Simulation (MASCS) [47] is an ABM developed to compare the effectiveness of several unmanned vehicle control architectures on carrier deck operations [46, 48]. Figure 2.7 is a snapshot of the MASCS simulation environment while running a 22 aircraft launch cycle. The simulation focuses mainly on the movement of the aircraft on the flight deck and the single operator type that guides this movement, which are the aircraft directors (shown as yellow dots in Figure 2.7).

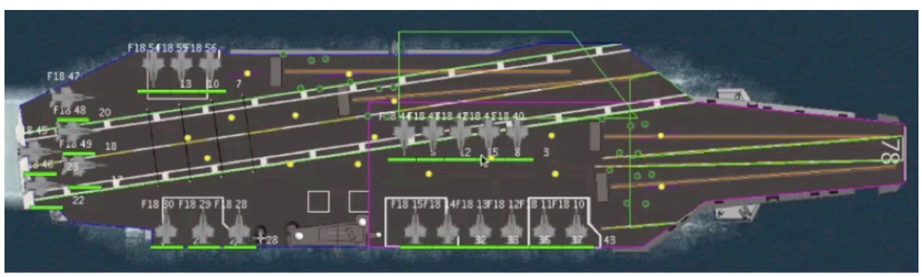

FIGURE 2.7: Original MASCS Simulator

The architecture of MASCS contains the four main components of an ABM outlined in [1] and described in Section 2.4. The model is built upon and visualized with GTGE [57] and written in the object oriented language of Java. The agents of the system include aircraft, aircraft directors and launch personnel, and catapults. The environment is a standard Nimitz-class aircraft carrier angled flight deck design with four catapults, the current design of all active US Naval carriers. [1]

The MASCS simulation does not include the behaviors of any squadron personnel, only deck operators. Nor does it include aircraft or catapult failures, which can severely affect the sortie rate. It does, though, demonstrate the ability of the agent-based approach to test scenarios based on the inclusion of new automation systems, such as automating the holdback bar installation at the catapults.

To further investigate the impact of adding automation to the flight deck as an integrated system (including aircraft, other deck vehicles and the various people that populate the deck), a new framework is needed. Chapter 3 describes in greater detail how the MASCS simulation can be extended to have the functionality needed to answer my research questions.

[1] The first carrier in the newest class of carrier deck designs, the Gerald R. Ford class, has just been built and is awaiting commissioning.

3

MASCS to PMASCS

This chapter describes the changes made to the original MASCS simulator in extending it to PMASCS and the major contributions of this thesis. Principally, this was the addition of aircraft failures and maintenance crewmen. Several additional changes were made including the introduction of aircraft-to-aircraft object avoidance and operator-to-aircraft object avoidance. Secondarily, four other squadron operator types were also added that are involved in the launch cycle but are not the focus of this thesis. All operators are discussed in Section 3.2.3.

3.1 MASCS

This thesis work builds upon an existing agent based simulation called the Multi-Agent Safety and Control Simulation (MASCS). MASCS is the result of a previous effort at the Massachusetts Institute of Technology funded by the Office of Naval Research.[1] Figure 2.7 is a snapshot of the MASCS simulation environment while running an aircraft launch cycle. The simulation was developed to compare the

[1] Integrating Global and Local Situational Awareness in Distributed Unmanned and Manned Ground Operations, Contract # N000140910625

effectiveness of several unmanned vehicle control architectures on carrier deck operations [47]. The simulation focuses mainly on the movement of one type of aircraft, the F18 fighter jet, on the flight deck and the single operator type that guides this movement, the aircraft directors (shown as yellow dots). Two other operator types were implemented in MASCS: the weightboard checkers and the holdback bar installers, both shown in green in Figure 2.7. The weightboard check and holdback bar installation occurs during the preparation of the aircraft for launch once it has reached the catapult.

MASCS measures a number of safety metrics, including primary, secondary, and tertiary halo incursions for the directors. Halo incursions are defined to occur when an agent comes within a certain range of a region determined to be dangerous, in this case proximity to a taxiing aircraft. However, the MASCS simulation does not include the behaviors of any other operators on deck: the ordnance operators, chocks and chains, fuel crews, safety officers, plane captains, and aircraft maintenance crews. Neither does it include a model of aircraft failures or other types of aircrafts. A new framework is needed to investigate the impacts on flight deck performance of operator skill training, maintenance operator staffing levels, and improved reliability of technology to aircraft. The next section describes in detail each of the extensions to MASCS which create the functionality needed to answer my research questions.

3.2 PMASCS

The following subsections describe the additions I have made in extending MASCS to Personnel Multi-Agent Safety and Control Simulation (PMASCS). I added four additional operator models and one additional aircraft type. The primary contribution of this work is two things: aircraft failures and maintenance crews. Figure 3.1 shows PMASCS at the beginning of a 22 aircraft launch cycle, with 15 F-18's, 7 E-2's, and a standard complement of operators.

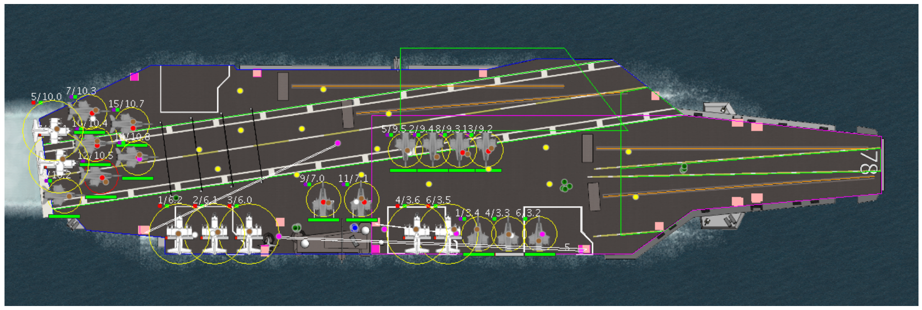

FIGURE 3.1: PMASCS Screenshot

3.2.1 Actions and Behaviors

All of the agents in the simulation have their behaviors defined as a series of actions to be completed. These actions and their associated distributions on rates, frequencies, or completion times are shown in Figure 3.2. Many of the actions are shared by every operator type, such as their walking speed and running speed. Some of the behaviors included here are not pertinent to the simulations in this thesis, but are nevertheless a part of PMASCS and included here for completeness. Specifically, the maintenance crews use a log normal distribution for their generic aircraft repair times. Due to the variability in human behavior and lack of carrier deck operator behavioral data, of particular difficulty is the determination and validation of appropriate task completion time distributions, which have been shown to be modeled well on gamma or lognormal distributions [8, 50]. We have chosen to use lognormal distributions in this model.

3.2.2 Aircraft

Each squadron on a carrier usually only has only one type of aircraft in it and a maintenance crew trained specifically for servicing that particular aircraft. On a normal carrier mission, there are usually 8 squadrons: 4 fighter squadrons, 2 helicopter squadrons, 1 E2C/D squadron, and 1 EA-18G squadron. Maintenance crews servicing F18 fighter jets would not normally be trained to work on E2 aircraft, or

Agent/Operator Type (Color)	Event/Task Affecting Operator Behaviors (Distribution Type)
Deck	landing zone faulted (exponential, λ = 0.10714), fuel station leak (exp., λ = 0.5/day), catapult failure rate (exp., λ = 7.716E-5)
Aircraft	maintenance needed (Bernoulli, p = 0.5, q = 1-p), taxi speed (Gaussian +/- 1SD, μ = 3.5, σ = 0.5), launch preparation time = (Gaussian +/- 1SD, μ = 109.648s, σ = 57.8s), takeoff acceleration (Gaussian +/- 1SD, μ = 2.76s, σ = 0.2717 s)
Operators - Generic	walking speed (Gauss., +/- 3.9SD, μ = 3.5mph , σ = 0.2mph), running speed (Gauss., +/- SD, μ =6.8 mph , σ = 0.2mph), fatality rate (exp., λ = 0.0062/yr), slip trip fall (exp., λ = 0.0015/person)
Checkers (Green)	weight board check time(Gauss. +/- 1SD, μ = 15s, σ = 5s),
A/C Maintainers (Green)	aircraft repair task time (LogNormal., μ = 6.8023s, σ = 0.0136s)
Operators (Green)	hold back bar installation (Gauss., +/- 1SD, μ =30s, σ = 10s), cat maintenance repair time (Gaussian, +/- 3.9SD, μ = 3500s, σ = 700s)
Grapes (Purple)	refueling speeds (Gauss., +/- 1SD, μ = 2000 lb/m , σ = 500 lb/m)
Chocks and Chains (Blue)	chocks removal time (Gauss., +/- 1SD, μ = 40s, σ = 10s)
Safety Observers (White)	Distribution on likelihood of a Safety Observer noticing a safety violation (e.g. fouled deck, fouled jet engine zone) given that the violation has occured. Prob(Observed|Violation = true)
Ordinance Officers (Red)	Ordinance Attachment (Gauss., +/- 2SD, μ = 20min, σ = 15), Ordinance Armament (Gauss., +/- 1SD, μ = 30s, σ = 10s)

Figure 3.2: Behavioral Distributions

any other aircraft for that matter, and visa versa. To investigate whether a generic training of maintenance crewmen would be beneficial to the overall launch efficiency of the carrier, I am including two types of aircraft and skill sets in my simulations. The two aircraft I have chosen are the generic F18 fighter jet and the E-2 Hawkeye. Specifications for each are shown in Table 3.1 and shown in Figure 3.3.

Table 3.1: Aircraft Specifications

Aircraft	Aircraft Fuel Capacity	Folded Wing Length	Unfolded Wing Length	Ordnance Attachment Time
F-18 Fighter (FMAC)	6780kg	8.38m	13.62m	150s (mean)
E-2 Hawkeye (SMAC)	5625kg	8.94m	24.6m	n/a

The aircraft also have many other specifications such as air speed, taxi speed, etc., but the above specifications are included here because they are pertinent to the behaviors I have developed for PMASCS, as described in the following sections.

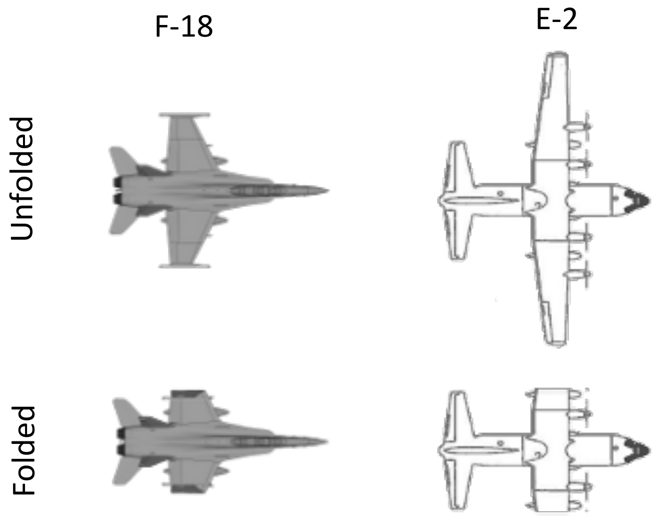

Figure 3.3: F-18 (FMAC) and E-2 (SMAC) Images

Aircraft-Aircraft Collision Detection

The original MASCS simulator included some basic object avoidance for the aircraft based on whether or not taxiing aircraft were in certain regions of the carrier and the catapults to which they were taxiing. It did not include any avoidance of operators, most likely because the aircraft Directors were directing the aircraft and all other operators were stationed at the catapults and not around the deck. Now with the inclusion of additional operators, it is necessary to include object and operator avoidance in the simulation because it can lead to additional delays that really do exist in launch operations.

The general approach is to have aircraft avoid other aircraft using a basic object avoidance technique but not to avoid operators, because operators are tasked with avoiding aircraft. When an aircraft is taxiing from its parking spot to a catapult, the path that it takes is split into a series of moves along smaller straight lines paths. There may be four or five of these paths in each taxi sequence, typically, and sometimes more if the catapult is far from the parking spot. On each of these moves, the aircraft checks for a collision continuously along the path of the current move action. This is achieved by projecting a polygon the width of half of the primary halo radius length along the straight line path of the aircraft. In Figure 3.4, this projected polygon can be seen. If the polygon intersects with another taxiing aircraft's primary halo boundary (right), the move action is paused until the path is clear. This generally prevents aircraft collisions. In the event that both aircraft are moving towards each other and both detect the other aircraft in a collision, each aircraft waits 60 seconds for the situation to resolve before moving forward, ignoring any future collision along that move path because it is likely to be the same aircraft with which it collided in the first place. This does not prevent aircraft from "running over each other" in the simulation, but it does resolve race conditions that would otherwise freeze the simulator, preventing it from finishing. The one-minute wait time is used to simulate the time lost in a real life scenario as the pilots and aircraft directors navigate the aircraft around one another.

Maintenance Failures

An aircraft may fly multiple missions each day. Prior to each launch, a series of maintenance checks is performed at various stages of preparation to ensure that there are no failures in need of repair. Repairs are so frequent and time consuming, relative to the amount of slack time an aircraft has between missions, that squadrons frequently keep one or two additional aircraft on hand and prepped for takeoff. In case of a failed aircraft, the mission cannot be postponed, so they will replace a failed aircraft with a functional one while maintenance is being completed if it will not finish in time. This is a costly resource, not only of time as any ordnances must

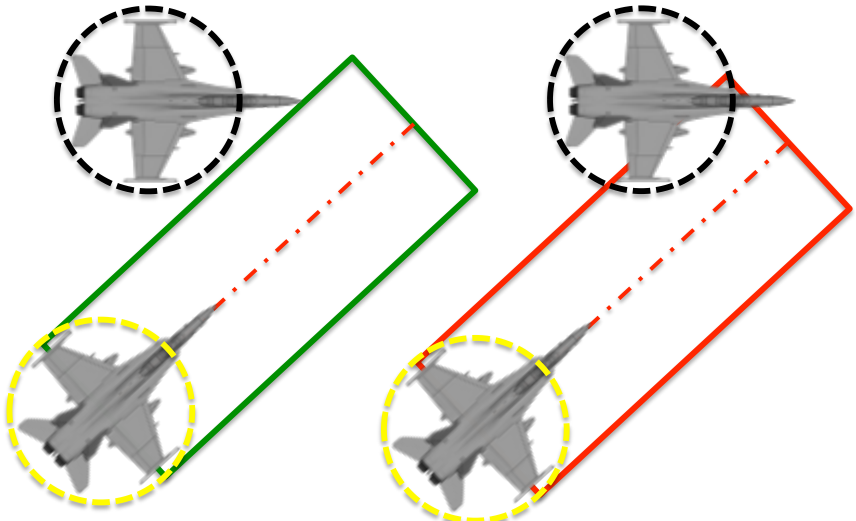

FIGURE 3.4: Aircraft-to-Aircraft Collision Detection

then be moved to the new aircraft, but also because of additional resources that are necessary to keep on deck. Space on the carrier deck is limited, so an additional two (potentially) unused aircraft per squadron is a use of valuable space. The focus of this thesis work is on the failures of aircrafts and repairs by the maintenance crewmen, towards identifying how this process can be optimized. To the author's knowledge, no other work on modeling of these failures and repairs of aircraft on aircraft carrier decks exists, and this is a novel contribution to the literature. In this section I discuss the failures of the aircrafts and their frequencies of occurrence. In a later section on maintenance operators, I discuss the repairs of the failures.

I have identified four phases of the launch cycle in which aircraft failures occur most frequently. Figure 3.5 graphically shows these four phases as:

- Phase 1: Prior to engine startup
- Phase 2: At engine startup, pre taxi
- Phase 3: During Taxi

- Phase 4: On the Catapult

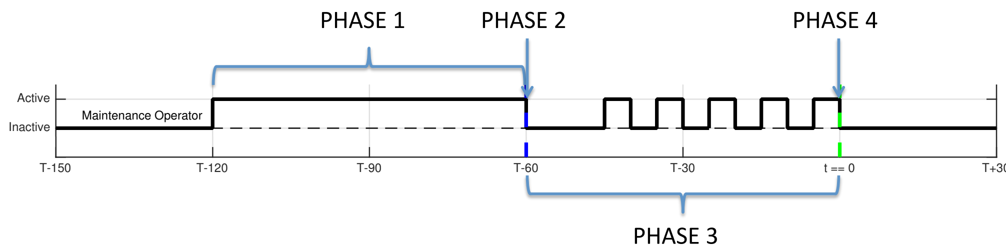

FIGURE 3.5: Launch Cycle Failure Phases

The black line indicates that maintenance crew are expected to be either active (high) or inactive (low) during that period. They are expected to be active early in the launch cycle and intermittently so during the rest of it.

Aircraft frequently have scheduled maintenance that must be performed. Since this maintenance is known about ahead of time, it is assumed that this maintenance is accounted for and finished in advanced such that it does not have the opportunity to affect launch time. So, it is not considered. The maintenance that is modeled here is for failures, i.e. required but unanticipated maintenance. That any aircraft can have no more than one maintenance failure per phase is an assumption made in this simulator. All failures are of a generic failure type, and the time to repair them is drawn from the same distribution each time. The likelihood of a failure in each phase is different, as well as the time of "discovery" or occurrence of those failures within the interval. Figure 3.2 shows the specifications of each phase of failures.

Table 3.2: Aircraft Failure Phases

Phase	μ, σ	Discovery Time (t), Distribution	p
Phase 1	900s, 150s	$t \leqslant 15min$, Uniform	0.5
Phase 2	900s, 150s	$t =$ Engine Startup, n/a	0.5
Phase 3	900s, 150s	$t \leqslant 30s$ after Start of Taxi, Uniform	0.15
Phase 4	900s, 150s	$t =$Arrival at Cat, n/a	0.05

The only change made to this distribution is the adjustment of the time it takes to repair a Phase 4 failure. This is discussed further in Chapter 4, but the purpose of this change is to see how the improvement of a commonly failed component on the aircraft that fails at launch time can improve the total launch times.

3.2.3 Operator Models

There are seven primary types of operators on the aircraft carrier deck relevant to aircraft maintenance and the launch cycle, each corresponding to a jersey color. They are explained in Table 3.3 based on observations, interviews, and Naval Operations manuals [38]. The role descriptions are not exhaustive, rather they describe the role of the respective operator for this simulation and their duties which pertains to the launch cycle.

Table 3.3: Operator Descriptions

Operator Type	Jersey Color	Standard Complement	Role Description
Aircraft Director	Yellow	11	Guide the aircraft during taxi to the appropriate catapult or parking spot
Safety Officer	White	6	Observe squadron operations and ensure safety protocols are followed
Ordnance Officer "ordies"	Red	9	Attach/detach and enable/disable ordnances to fighter aircraft
Equipment Operator / Chocks and Chains	Blue	24	Maneuver the tractors and (un)chock aircraft
Maintenance and Catapult Crew	Green	20	Perform schedule and unscheduled maintenance on aircraft; perform weight checks and install holdback bar
Fueling Operator "grapes"	Purple	18	Refuel aircraft prior to launch
Aircraft Captain	Brown	22	Ensure all fueling, maintenance, ordnance, and chocks and chains are removed prior to taxi

The original MASCS simulator modeled the Aircraft Directors and Catapult Crew. The following sections describe the operator types that are now implemented in PMASCS. Of primary focus for this thesis was the maintenance crewmen, so extra attention is given to them.

Ordnance

Ordnance officers are tasked with attaching ordnances to the aircraft during the preparation phase of the launch cycle. They wear a red jersey and are shown as red dots in Figure 3.1. It is not uncommon for maintenance crew to be active during the taxi phase and on the catapult, because ordnances are not engaged at the time of attachment. They are engaged just before take off. This behavior is not implemented in PMASCS, and it is assumed that the time to complete a Phase 3 failure encompasses this action well enough. All F-18 aircraft receive an ordnance upon the beginning of the launch cycle, but E-2's do not receive any.

Maintenance

Maintenance operators are responsible for all preparatory and failure related maintenance on aircraft and catapults. They wear green jerseys and are shown as green dots in Figure 3.1. There are four phases of maintenance for which the maintenance operators are responsible, as shown in Table 3.3. Since maintenance on aircraft takes much more time than any of the other preparatory procedures, the amount of time spent on maintenance has a direct impact on the total time of the launch cycle. Each squadron has its own maintenance crews specifically trained to maintain the squadron's specific aircraft. F-18 squadron maintenance crewmen can only work on F-18s and would generally only work on their specific squadron's F-18's, similarly so for E-2 squadrons. As will be discussed further in the Analysis and Results chapter, I have created three types of skill sets for the maintenance crew. They may either be trained to maintain E-2's, F-18's, or have a generic skill set allowing them to maintain both aircraft.

The operator first determines whether they have a generic skill set or a specific skill set, and then chooses an aircraft from a list of aircraft needing maintenance. They then walk across the deck to that aircraft and draw from a lognormal distribution to determine how much time will be required to fix the aircraft. Distributions for this are shown in Table 3.2. Once they are complete with the maintenance, they return to the tower, simulating the need to return tools and get parts for their next assignment. If there are no aircraft needing service, the maintenance operators return to a safe zone in front of the tower to wait.

Plane Captain

Plane Captains are responsible for ensuring that all checks prior to taxi are complete and the aircraft is ready for take-off. They are the gate keepers for the movement of the aircraft, and are shown as brown dots in Figure 3.1. Plane Captains ensure that the aircraft is fueled, maintenance is completed, ordnances are attached, there are no failures on the aircraft, and that the aircraft is unchocked before it is marked as ready to taxi. The Plane Captains stay with the aircraft until it is unchocked, and then return to a safe home place near the tower once their aircraft is marked as ready.

Chocks and Chains

The Chocks and Chains crew move in pairs and are responsible for unchaining and unchocking the aircraft prior to taxi. If an aircraft begins taxiing but must wait for an extended period of time, it is rechocked. Delays can occur if there is a delay on the flight deck due to a failed catapult or the carrier is making a turn. In PMASCS, the chocks and chains crewmen only unchock the aircraft once and then return to a home zone near the tower. Each team of two are represented with blue dots in Figure 3.1.

Safety Officers

The Safety Officers are operators on deck with each squadron whose sole responsibility is to ensure that the other operators are following safety protocols. They keep watch over the launch maintenance operations, calling out operators if they walk into a zone they are not supposed to such as in the landing zone, causing a faulted deck, or in front of a running engine. The Safety Officers are present in PMASCS, as shown in Figure 3.1 as white dots, but do not have any influence on the launch cycle. They move from aircraft to aircraft in each of their respective parking spots and "observe" but do not interfere with the maintenance operations. It is conceivable that the Safety Officers would notice a safety violation with the maintenance operators and delay their work, but this is a rare event and considered marginal in the overall launch times. So, it is not considered in this thesis.

Operator-Aircraft Collision Detection

Similar to the aircraft-aircraft collision detection described above, the operators have operator-aircraft collision detection. Aircraft do not avoid operators; it is the operators' job to avoid taxiing aircraft. When an operator moves across the deck, they take the shortest path distance. In reality, operators would take the shortest path distance while avoiding danger areas, such as walking directly in front of a moving aircraft or running engine which could put them and the plane in danger. This behavior is not yet implemented in the simulation, except that they will wait to move if there is a taxiing aircraft in their projected path. This helps to reduce the number of halo violations, one metric for safety of the operators. Figure 3.6 shows the projected path of a person as a polygon the width of the dot representing a person and along the path of travel. If the polygon intersects a taxiing aircraft's primary halo radius, the operator waits in place until the collision is no longer detected. No timeout has been implemented in this scenario. Since aircraft do not avoid operators, the gridlock cannot occur and it is not necessary to implement this. An aircraft is not considered taxiing if it has not moved in the last time step of the simulation, so an aircraft in the Action of Taxi awaiting another aircraft is not considered a taxiing aircraft for the sake of operator-aircraft collision detection.

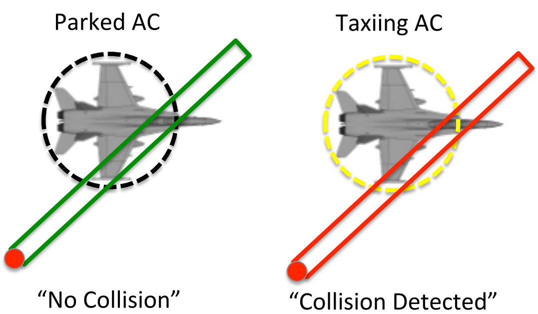

FIGURE 3.6: Operator Object Avoidance

In this chapter, I have discussed all of the new changes made to the original MASCS simulator, extending it to PMASCS. Chapter 4 discusses the experimental conditions that were tested in PMASCS and analyzes the results to answer the hypothesis put forth in the introduction.

4

Simulations, Results, and Analysis

Towards the optimization of maintenance crew manning on aircraft carrier decks, I've designed four experiments to test my hypotheses. The first experiment, Experiment A, simulates current manning conditions across various manning levels from 1 team to 15 teams. Experiment A is akin to the normal carrier deck conditions and is used to verify that PMASCS correctly captures the dynamics of a launch cycle. I then perform three experiments with changes to either skill set training of maintenance crew, improved technologies allowing for fast repairs during Phase 4 failures, or a combination of both. Figure 4.1 shows the combinations of the experimental conditions and their associated experiment labels. Table 4.1 explains each of the experiments and the maintenance levels tested. For each setting, I run the simulation for various numbers of manning levels so that a minimal manning level, at least, can be determined from the average total launch times under each experimental conditions. From there, a subject matter expert could determine the safety factor of manning desired on the carrier deck, and back track to a desirable level for optimal manning, balancing financial objectives with preparedness objectives using actual workload estimates and measures of performance.

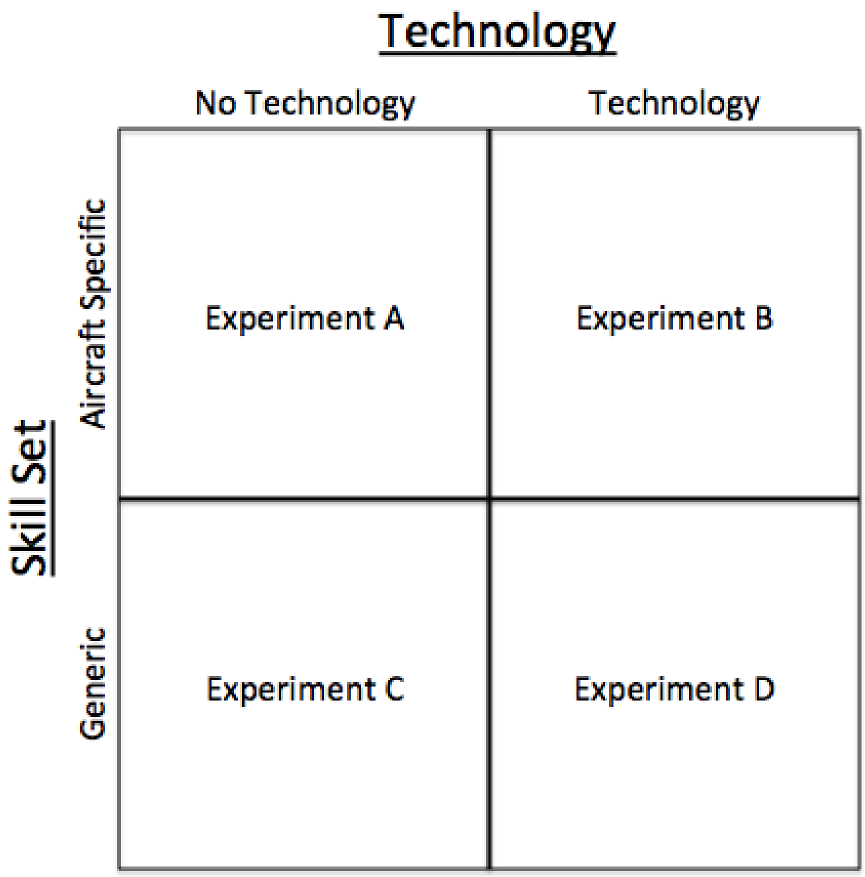

Figure 4.1: Combination of Experimental Conditions

The first experiment, Experiment A, determines a baseline for the simulator and launch times. The total launch time was validated by SMEs and observations on aircraft carrier decks. This baseline is used as a basis for comparison for the other experiments. Experiment B explores how a technology improvement to the last stage of failures, Phase 4, which occur on the catapult, can be hastened to improve launch rates. In current operations, if an aircraft fails on the catapult and it is reparable quickly, it is repaired on the catapult, preventing the use of that catapult by other planes. This can severely delay launch operations of aircraft in that catapult's queue. So, Experiment B explores any possible improvements that can be gained from fixing a maintenance failure with an improved technology that is quicker to fix. Experiment

Table 4.1: Experiment Descriptions and Manning Levels

Experiment Name	Description	Manning Levels
Experiment A	Maintenance operators are trained in either E-2 or F-18 Maintenance and must work only on their specific aircraft. Phase 4 failures require the normal amount of time to fix.	2 - 15 Teams Total, 1 - 10 F-18, 1-5 E-2
Experiment B	A technology improvement is made to decrease the completion time of Phase 4 failure maintenance by a factor of 10. Maintenance operators are trained in either E-2 or F-18 Maintenance and must work only on their specific aircrafts.	1 - 15 Teams
Experiment C	Maintenance operators are trained on both E-2 and F-18 aircraft and may work on either. Phase 4 failures require the normal amount of time to fix.	1 - 15 Teams
Experiment D	Maintenance operators are trained in either E-2 or F-18 Maintenance and must work only on their specific aircrafts AND a technology improvement is made to decrease the completion time of Phase 4 failure maintenance by a factor of 10.	2 - 15 Teams Total, 1 - 10 F-18, 1-5 E-2

C examines how relaxing the constraints on maintenance crewmen to working on one specific aircraft can have an affect on the total launch times. It is expected that launch times would improve, as state in Hypothesis 1B, since this is a looser constraint than having plane-specific skill sets since any available operator can work an any plane in need of maintenance. Experiment D combines both the technology improvement and a generic skill set training to see how it would affect launch times. The Kruskal-Wallis H (KWH) test is used to show statistical significance between manning levels and across experimental conditions.

The following sections describe the experiments in more detail and present results from each, including an example timeline for launching aircraft under that experiment and a KWH analysis to show statistical significance. Pairwise comparisons are used to show statistical significance for each of the manning levels within a set and to identify the critical manning level corresponding to the "elbow" in trade-offs between additional operators and improved total launch time. Finally, a Pareto analysis is presented to compare the experimental results and determine an optimal global manning level and condition.

4.1 Experiment A (Skill Separation / No Technology Improvement)

Each squadron generally has one aircraft type and maintenance squadrons that are appropriately trained for that aircraft. This is the baseline condition for normal carrier deck operations, the one against the other two simulations are compared. The number of F-18 maintenance teams was varied from 1 to 10, while the number of E-2 maintenance teams was varied from 1 to 5 based on the Equation 4.1, where $numE2Teams$ and $numF18Teams$ denote the number of E-2 maintenance teams and number of F-18 maintenance teams, respectively, and $ceil()$ represents the ceiling function, which rounds up to the nearest whole number.

$$numE2Teams = ceil((7/15) * numF18Teams) \tag{4.1}$$

This adds roughly one E-2 team for every two F-18 maintenance teams, on average. This relationship was derived from SME knowledge that for every 4 F-18 on deck there are 4 maintenance crewmen, and for every E-2 on deck there are 9 crewmen. This is an average of 4 crewmen per F-18 and an average of 4.5 crewmen for every E-2. Because maintenance crews work in teams, and for simplicity of the simulation, I have represented the teams as units that can service one aircraft at a time. So, the maintenance teams (a.k.a. the dots in the simulation) represent a four-man maintenance crew if it working on an F-18 and an average of a 4.5 man maintenance crew if it is working on an E-2.

Table 4.2: Failure Phase Descriptions

Failure Phase	Descriptions	Probability of Occurrence
Phase 1	Phase 1 failures are discovered prior to engine startup through maintenance checks, and may be discovered anytime during the first 15 minutes of aircraft checks.	$p = 0.5$
Phase 2	Phase 2 failures are discovered immediately upon startup of the engine.	$p = 0.5$
Phase 3	Phase 3 failures occur while an aircraft is taxiing to a catapult and can occur anytime during the taxi.	$p = 0.15$
Phase 4	Phase 4 failures may only occur on the catapult and are discovered immediately upon reaching the catapult.	$p = 0.05$

Scheduled maintenance on aircraft is started nearly two hours prior to the launch of the first aircraft of a sortie. The focus of these experiments is not on scheduled maintenance, rather on unexpected maintenance failures. This simulation begins 45 minutes prior to the start of taxi of the first aircraft. Table 4.2 describes each type of failure that may occur and their likelihood of occurrence. Phase 1 failures may occur anytime in the first 15 minutes of the simulation. At 15 minutes, or whenever the Phase 1 failure is complete if there is one on the aircraft, the engine is started. At engine startup, Phase 2 failures may be discovered. At no earlier than 30 minutes, after all failures have been repaired, an aircraft may begin taxi. Often times aircraft are queued in the taxi and awaiting a free catapult. Catapults allow up to three queued aircraft at a time that have taxied from their parking spots. Anytime during the taxi, an aircraft may have a Phase 3 failure. During a Phase 3 failure, the aircraft stays in place and is repaired. Once the taxi is complete and the aircraft is on the catapult, a Phase 4 failure may occur. If it occurs, the aircraft is held on the catapult until the failure can be fixed and it is then launch.

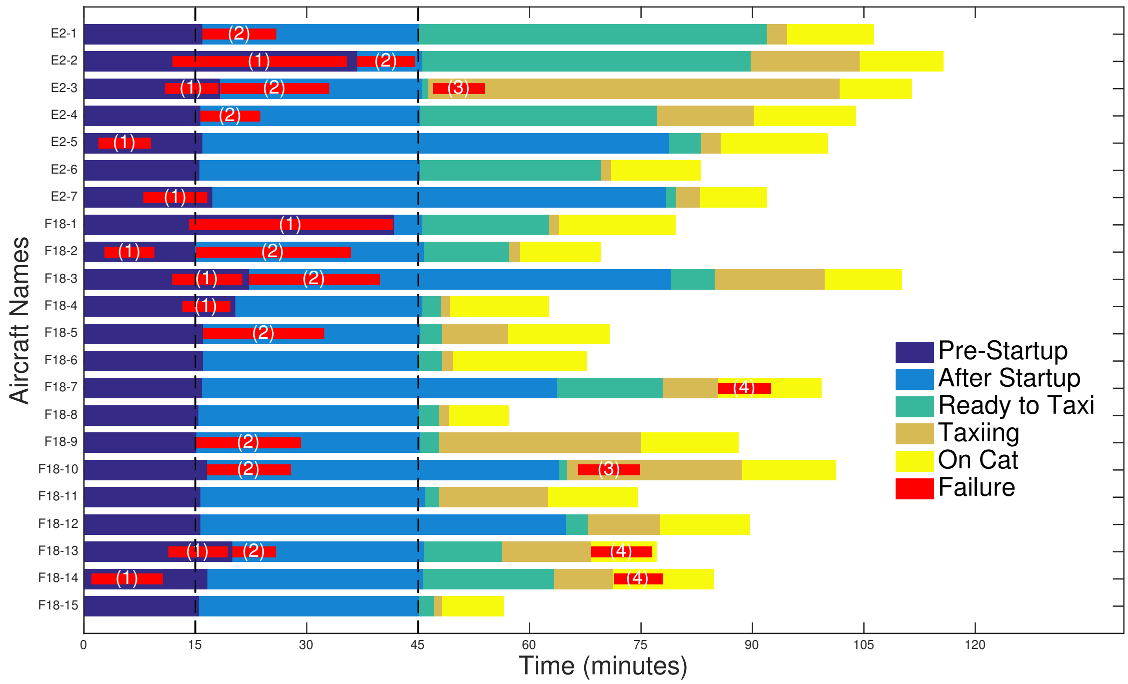

FIGURE 4.2: Timeline of Launch Cycle for ExperimentA

Each aircraft attempts to follow a standard schedule for maintenance, taxi, and launch. Figure 4.2 shows an example timeline for each of the 22 aircraft being launch in a generic launch cycle. Since maintenance may result in delays as well as readied aircraft queuing behind catapults leading to delays in the start of taxi, each aircraft's launch schedule differs from the next. Failures requiring maintenance are shown in red bars superimposed on the timeline. The numbers in each red back indicate which phase the failure occurred in. The total launch time is the time from the beginning of the simulation until the last aircraft is launched. In Figure 4.2, this is aircraft E2-2. Figure 4.3 and Table 4.3 show the minimum, maximum, mean, median, number of runs, and standard error for total launch times at each of the manning levels for 22 aircraft with maintenance crews with specific skill sets. Dashes, '-', represent manning levels that were not tested because that manning level is not achievable with the way the number of maintenance crews is incremented per equation 4.1.

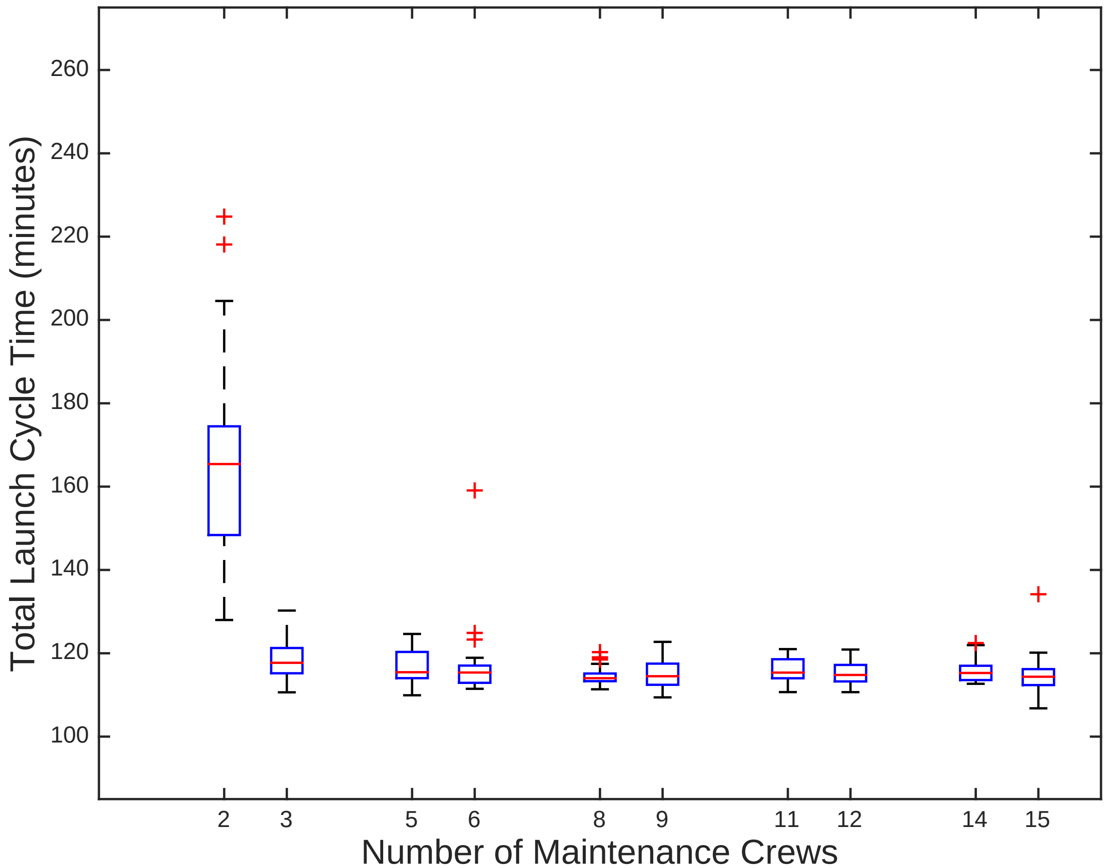

FIGURE 4.3: Total Launch Times of Experiment A, 7 F-18 Maintenance Crews, 3 E-2 Maintenance Crews

4.1.1 KHL Test

It is clear from these results that the average launch time is between 115 and 118 minutes for any manning level above 3 teams, with no noticeable trend or improvement for additional maintenance teams. A KWH test showed that there was a statistical difference between the manning levels with $p = 3.088e-14$. Figure 7.1 in Appendix B shows the shape of the distributions for each of the manning levels. The left most

Table 4.3: Total Launch Time Distribution Data for Experiment A

Level	Num. Runs	Minimum	Maximum	Mean	Median	Standard Error
1	-	-	-	-	-	-
2	26	127.9882	224.8128	167.3473	165.4175	4.9285
3	27	110.6292	130.2518	118.4420	117.7115	0.9410
4	-	-	-	-	-	-
5	24	109.9210	124.6365	116.7785	115.4532	0.7987
6	23	111.4710	159.0840	117.4069	115.3767	2.0215
7	-	-	-	-	-	-
8	23	111.3578	120.2747	114.6833	114.0162	0.4863
9	26	109.4155	122.7362	115.0344	114.4920	0.6664
10	-	-	-	-	-	-
11	27	110.6818	120.9983	115.8755	115.3577	0.5245
12	27	110.6697	120.8978	115.2095	114.7948	0.5368
13	-	-	-	-	-	-
14	27	112.6768	122.4523	115.7217	115.2708	0.4938
15	22	106.7880	134.1685	115.1445	114.3684	1.1197

on the top row is for manning level 2, and the right most on the bottom row is for manning level 15. The missing plots are for manning levels that are not achievable with skill separation because of 4.1.

Per assumption 4 of the KHL test, the distribution "shape" is similar for all except the manning level 2, so the KHL test can only be used to compare mean ranks. Results from the KHL pairwise comparisons test are shown in Appendix B, Figure 4.4 with alpha value set using a Bonferroni correction for family wise error. The raw data in table format is shown in Chapter 6.

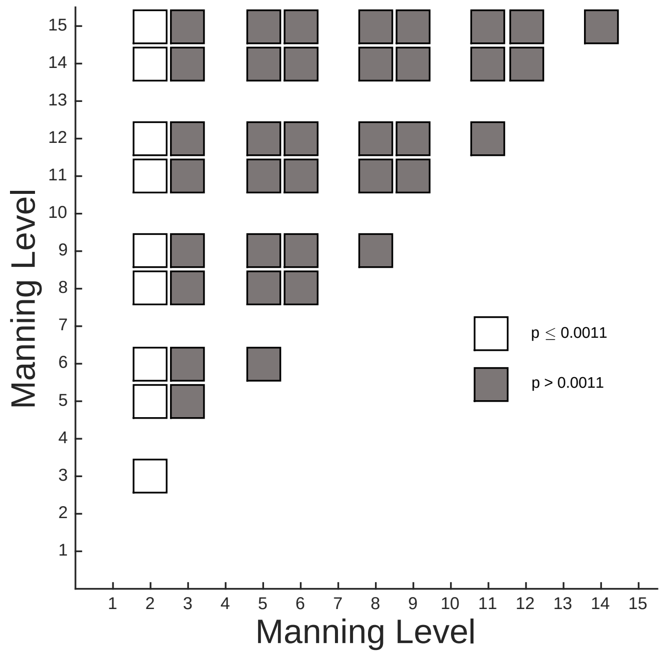

FIGURE 4.4: Pairwise Comparisons for Experiment A

The KHL test is showing that the mean ranks of launch time distributions are not statistically significantly different for any of the manning levels 3 to 15, but that manning level 2 is statistically significantly different. This means that there is no statistically significant decrement in the mean total launch time until the manning levels reach 2, which is one E-2 maintenance crew and one F-18 maintenance crew. In other words long as there is sufficient crew members to service one E-2 and one F-18 simultaneously, under skill separation, the normal launch rate is attainable with no decrement to the total launch time. This I will refer to the "elbow point" of the curve. It can be seen in Figure 4.3 that there is an "elbow" in the total launch times between manning levels 2 and 3, corresponding to the results from the KWH test. The location of this elbow changes depending on the training level of the maintenance crews as well as the time required to fix Phase 4 failures, as can be seen in the following two experiments.

4.2 Experiment B (Technology Improvement & Skill Separation)

The catapult is the bottleneck for launch operations, since there are a limited number of them and many fewer than the number of aircraft that need to be launched typically. So, if a catapult is either failed or effectively down because of an aircraft failure on it, preventing other aircraft from launching, the efficiency of the launch cycle can be severely affected. Aircraft are allowed to have a Phase 4 failure once they are on the catapult. In the previous two experiments, this Phase 4 failure was the same type of generic failure as in any other case. In this experiment, I speed up the Phase 4 failure maintenance time by 10x the original, i.e. it takes 10x fewer minutes to fix the failure, on average, than normal. It is conceivable that an improvement in technology could achieve this. Figure 4.5 shows a boxplot of the total launch times for manning level seven. Similar to Experiment A, there is not a significant change in the launch times until manning level two. Table 4.4 shows the numerical results from this simulation.

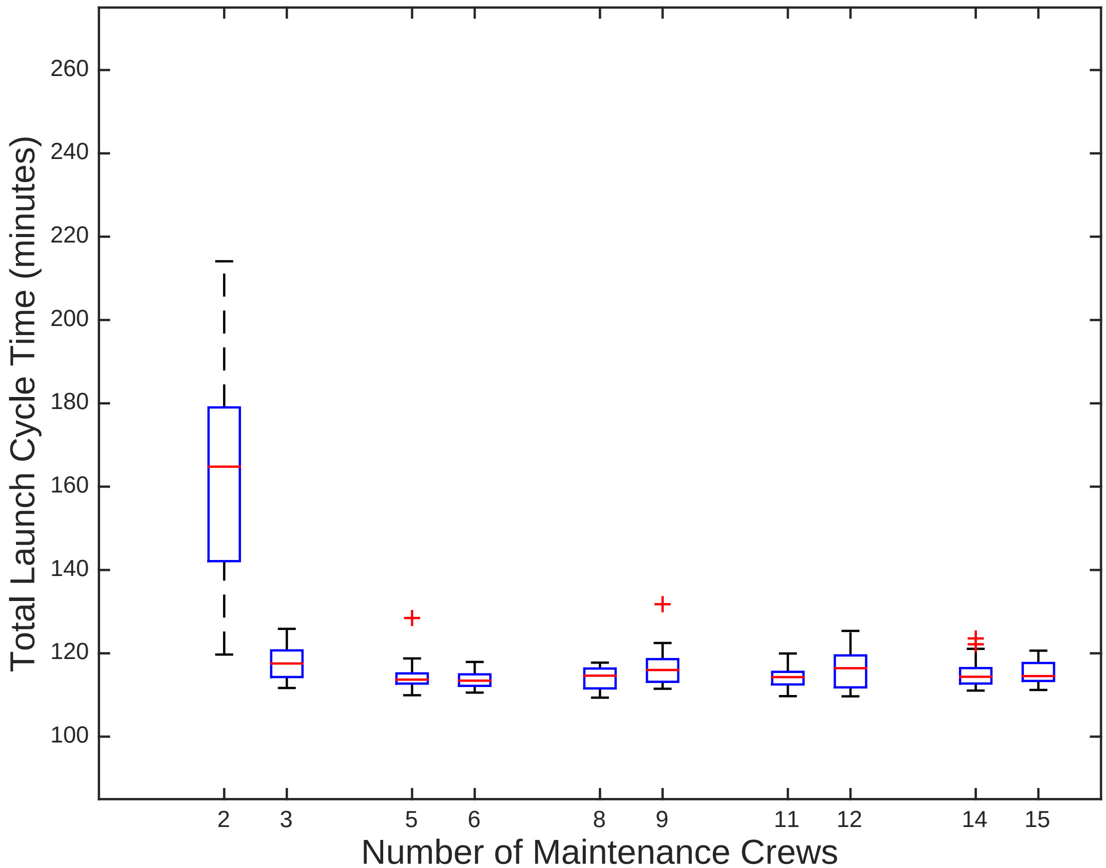

FIGURE 4.5: Total Launch Times of Experiment B, 7 Generic Maintenance Crews

4.2.1 KHL Test

The average total launch times for between manning levels 3 and 15 are all fall between 113 minutes and 118 minutes, however at manning level 2 the average total launch time is much greater. A KWH test showed that there was a statistical difference within the manning levels with $p = 1.279e - 13$. A pairwise comparison test is used to show that manning level 2 is statistically different than the other manning levels. In Figure 7.2 in Appendix B the distributions of launch times vs manning level are shown.

Figure 7.2 shows the distributions at each manning level to be all similar except

Table 4.4: Total Launch Time Distribution Data for Experiment B

Level	Num. Runs	Minimum	Maximum	Mean	Median	Standard Error
1	0	-	-	-	-	-
2	24	119.703	214.093	160.501	164.797	1.011
3	24	111.678	125.871	117.786	117.554	0.172
4	0	-	-	-	-	-
5	26	109.936	128.477	114.308	113.687	0.135
6	26	110.569	117.928	113.885	113.438	0.077
7	0	-	-	-	-	-
8	25	109.358	117.761	114.110	114.635	0.107
9	24	111.492	131.785	116.648	115.993	0.186
10	0	-	-	-	-	-
11	25	109.717	119.958	114.431	114.289	0.096
12	25	109.675	125.376	115.962	116.423	0.169
13	0	-	-	-	-	-
14	27	111.054	123.569	115.154	114.368	0.121
15	22	111.206	120.637	115.276	114.536	0.123

for manning level 2, so we can only use the KHL test to compare mean ranks for statistical significance.

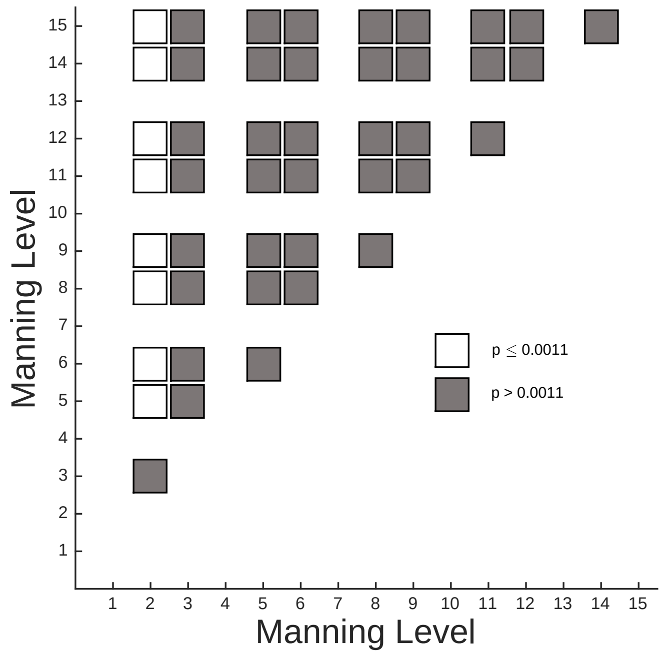

FIGURE 4.6: Pairwise Comparisions for Experiment B

Figure 4.6 shows the results from a pairwise comparison tests of all the combinations of manning levels. Not surprisingly, manning level 2 is significantly different than most all of the other manning levels. From this analysis, the minimal manning level with a technology improvement and skill-separated crew, is manning level three. The next section looks at how skill levels of maintenance operators can affect total launch times.

4.3 Experiment C (Generic Maintenance Skill Set)

Maintenance crews are usually constrained to their squadron and by their specific training, allowing them to only work on one aircraft. This experiment is designed to identify how training the maintenance crews with a generic skill set will affect the total launch time of sorties. In my hypothesis, I expected that the launch times would decrease with a generic skill set because if there is an available maintenance operator for a failed aircraft, it can be serviced, rather than the aircraft having to wait for a specific type of maintenance operator to service it. Figure 4.7 and Table 4.5 show the minimum, maximum, mean, median, number of runs, and standard error for each of the manning levels.

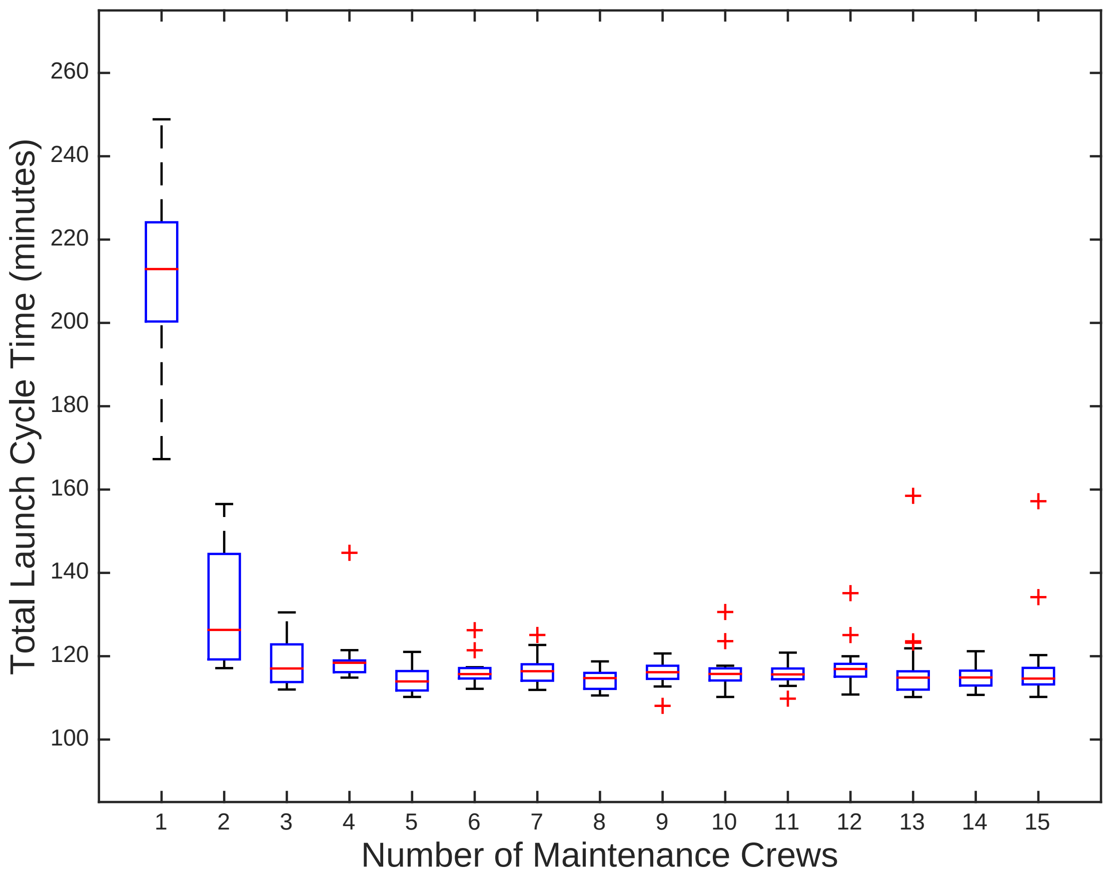

FIGURE 4.7: Total Launch Times for Experiment C, 7 Generic Maintenance Crews

4.3.1 KHL Test

A KWH test showed that there was a statistical difference between the manning levels with $p = 3.492e-11$. For manning levels 4 through 15, the mean is nearly the same, between 114 minutes and 116 minutes, with little or no improvement for additional maintenance teams. Manning levels 2 and 1, however, do show a significant change in the mean total launch times from the rest of the manning levels and from each

Table 4.5: Total Launch Time Distribution Data for Experiment C

Level	Num. Runs	Minimum	Maximum	Mean	Median	Standard Error
1	11	167.3075	248.8643	212.3839	212.9208	6.9632
2	13	117.1377	156.5263	132.1964	126.3138	3.9922
3	12	112.0058	130.5160	118.7654	117.0559	1.8941
4	12	114.8607	144.8175	119.9303	118.4193	2.3305
5	14	110.2245	121.0323	114.2587	113.9470	0.7950
6	13	112.1760	126.2445	116.6304	115.7085	1.0061
7	26	111.9132	125.0965	116.3753	116.4001	0.6303
8	22	110.5768	118.7538	114.4472	114.7463	0.5111
9	23	108.0860	120.6578	115.8887	116.1580	0.5683
10	15	110.2205	130.6112	116.5826	115.7410	5.9629
11	26	109.8127	120.8595	115.7893	115.6306	0.4940
12	24	110.8022	135.1317	117.4369	116.9293	0.9494
13	23	110.2000	158.4965	116.8799	114.8607	2.0513
14	25	110.7157	121.1938	115.0408	114.8973	0.5270
15	29	110.2253	157.2023	116.8458	114.6333	1.6564

other, as can also be seen in Figure 4.7. Figure 7.3 in Appendix B shows the shape of the distributions for each of the manning levels.

Except for manning levels 1 and 2, the distributions are similar in shape. Because of Assumption 4 of the KHL test, it can be used only to compare the mean ranks of the distributions for statistical significance. Results from the KHL pairwise comparison tests are shown in Figure 4.8 with alpha value set using Bonferroni correction for family wise error. The raw data in table format is shown in Appendix A.

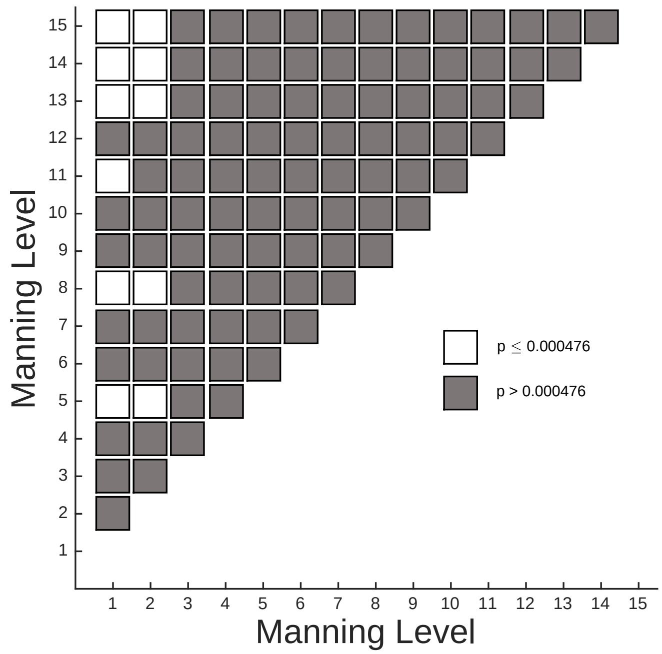

FIGURE 4.8: Pairwise Comparisions for Experiment C

The pairwise comparison test is showing that the mean ranks of launch time distributions are not statistically significantly different for any of the manning levels 3 to 15, but that manning levels 2 and 1 are statistically significantly different. The elbow, then, for this plot is also at manning level 2, the same location of the elbow in the Skill Separation case. This is visible in Figure 4.7. This, in effect, nullifies Hypothesis 2a that a generic skill level will be allow for a small complement of maintenance crews over the standard complement.

4.4 Experiment D (Generic Skills & Improved Technology)

In Experiments B & C, a technology was introduced to improvement maintenance times and restriction was relaxed through general skill set trainings, respectively. In this experiment, both are implemented to see how making both changes to carrier deck operations will affect total launch time. Figure 4.9 and Table 4.6 show the results from this simulation. Similar to the Generic Skill set results, the elbow of the curve is near manning level 2.

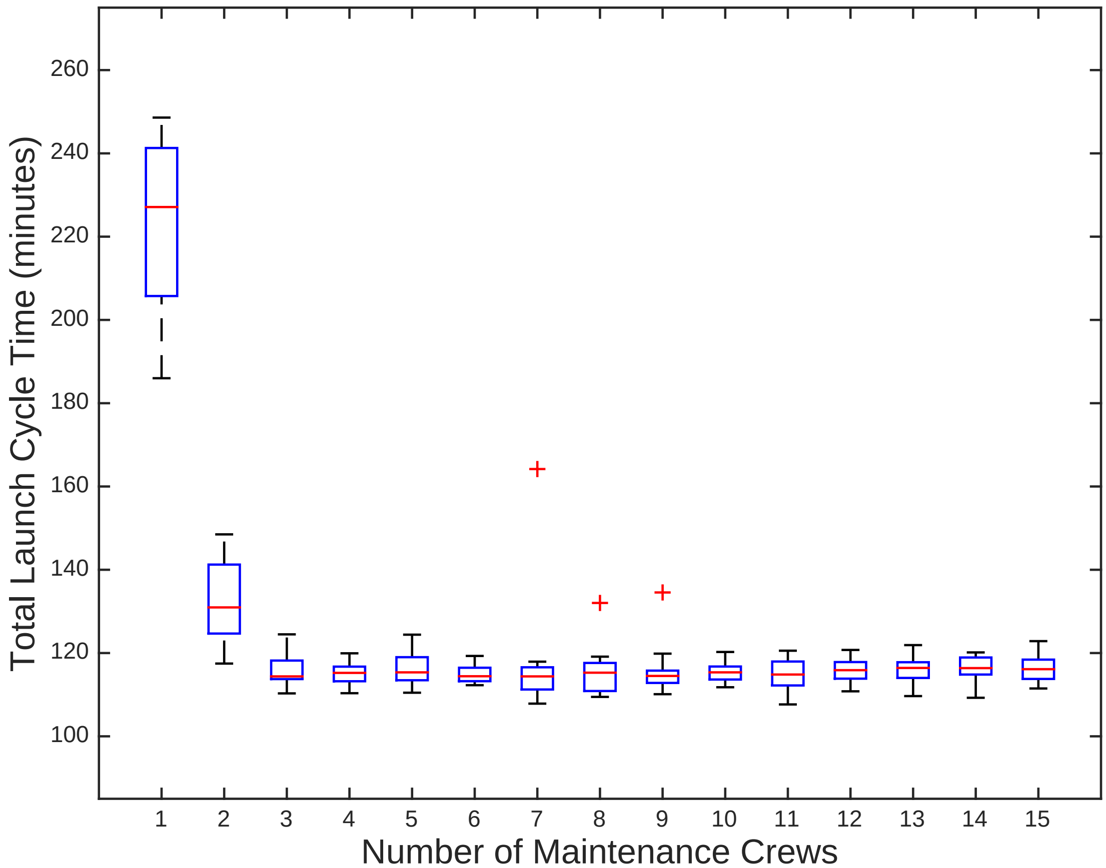

FIGURE 4.9: Total Launch Times of Launch Cycle for Improved Technology on Phase 4 Failures

4.4.1 KHL Test

The mean total launch time for manning levels 3 through 15 has a very low variance, and is bounded by 114 minutes and 117 minutes. Consistent with the other experiments, manning levels 2 and 1 are significantly different, as can been seen in the means of the distributions in Table 4.6. A KWH test showed that there was a statistical difference between the manning levels with $p = 1.34e - 17$. Figure 7.4 in Appendix B shows the shape of the distributions for each of the manning levels. The pairwise comparison results are shown in in Figure 4.10.

Table 4.6: Total Launch Time Distribution Data for Experiment D

Level	Num. Runs	Minimum	Maximum	Mean	Median	Standard Error
1	23	185.9992	248.5955	222.2255	227.1113	4.3486
2	18	117.4725	148.4973	132.0233	130.9587	2.2655
3	15	110.3088	124.4868	116.0951	114.3888	1.0545
4	15	110.3553	119.9345	114.7740	115.2340	0.6539
5	16	110.4725	124.4108	116.0417	115.3605	0.9679
6	15	112.2790	119.3020	114.9324	114.4283	0.5412
7	17	107.8507	164.1877	116.6131	114.3905	3.0624
8	18	109.4650	132.0288	115.2210	115.2798	1.2424
9	26	110.1167	134.5493	114.8816	114.5002	0.8917
10	26	111.7843	120.2555	115.2499	115.3615	0.4102
11	24	107.6573	120.5610	115.0170	115.0170	0.7049
12	25	110.8078	120.7425	115.9384	115.8835	0.5638
13	24	109.6615	121.9052	116.1166	116.4205	0.6126
14	26	109.2572	120.1497	116.1865	116.3816	0.5672
15	23	111.4882	122.8598	116.2557	116.1240	0.6196

Even more clearly than in the Generic Maintenance results, the pairwise comparison shows a statistically significant change at the elbow and for manning level 1. The Phase 4 failures, then, seem to not have a large impact on the mean launch rates since these results differ little from the Skill Separation and Generic Skill experiments. I will show, though, in Section 4.5 that it may still improve launch rates from the perspective of a lowest-achievable total launch time.

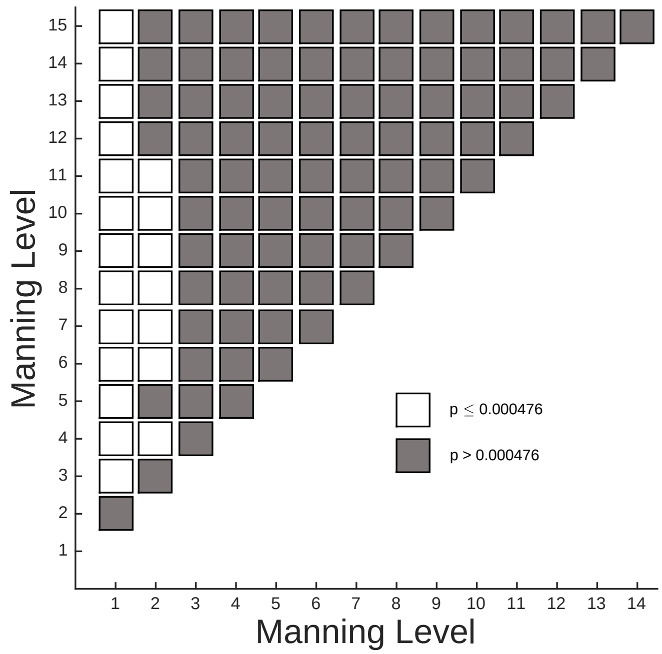

FIGURE 4.10: Pairwise Comparisons for Experiment D

4.5 Pareto Analysis

A Pareto analysis is a multi-objective optimization that can be used to determine the "optimal" tradeoff between two conflicting objectives, commonly cost vs. performance [36]. In the case of the carrier deck, not only is it more financially costly but it also increases the risk to operators the more of them that are on deck. It can be argued that the fewer people in harms way means there are fewer people to get injured means an overall safer deck environment. Cost savings, especially on the carrier deck, should not be made at a significant cost to performance or safety, because ship readiness is necessary for the military. So, the question at hand is how to balance costs of maintenance crew's wages and safety vs. deck performance. As posed in the introduction, one of the research question of this work is "What is the fewest number of maintenance crews allowable to maintain the nominal launch rate?" This question is answerable using a Pareto frontier analysis, show in Figure 4.11.

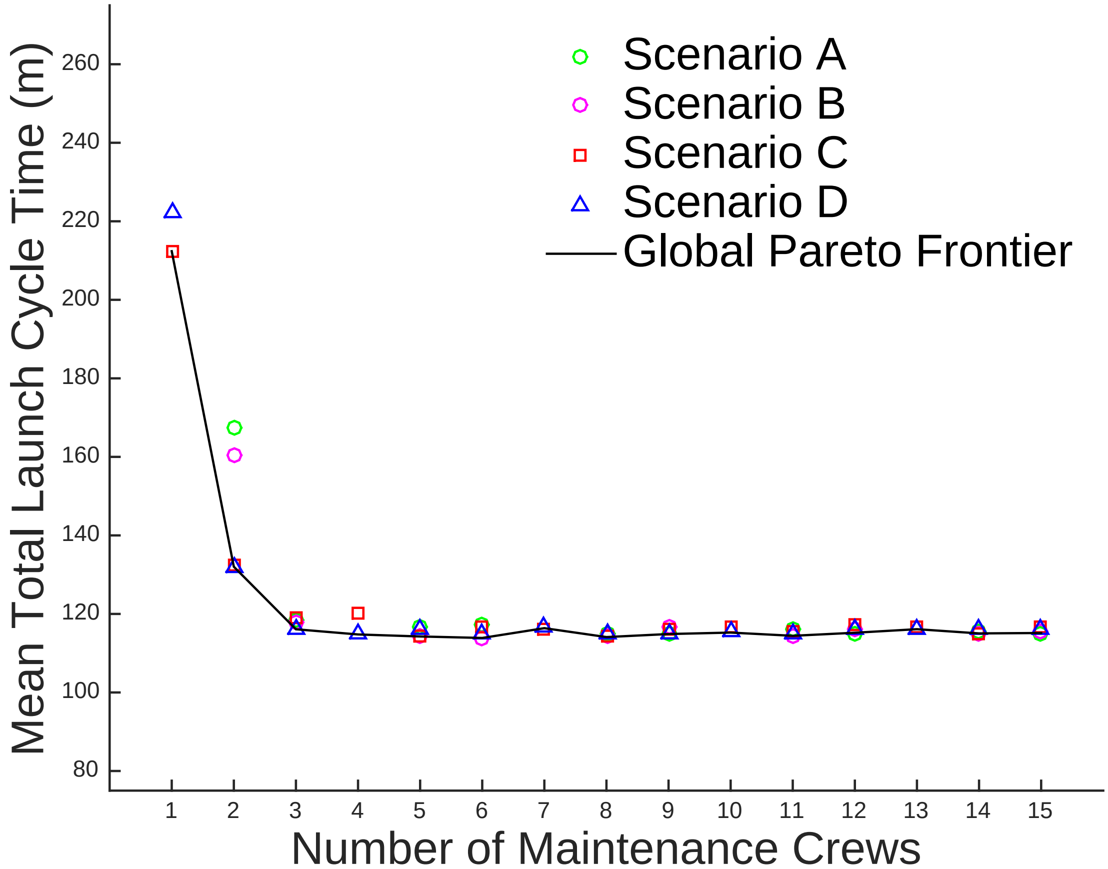

FIGURE 4.11: Pareto Frontier across Manning Levels for Launch Time

For each of the four experiments, Figure 4.11 shows as a scatter plot of the means of the total launch times for each simulation at each manning level. The black line hugging the bottom of the scatter plots is the Pareto frontier, and marks the "limit" of the possibilities of outcomes, in this case the minimum of the four experimental means. By moving to the right or left on the frontier, you can determine if there is a trade off in one or other of your constraints. A flat (or vertical) line on the frontier means that there is no trade off in the competing constraint whilst moving along that portion. The frontier for all four experiments from 15 down to 3 manning levels is approximately flat, i.e. there is no clear trade off in terms of the competing constraint (y-axis) of total launch time. In other words, you do not lose any performance in

Table 4.7: Mean Launch Times Comparison Across Experiments and Manning Levels

Level	A	B	C	D
1	-	-	212.3839	222.2255
2	167.3473	160.501	132.1964	132.0233
3	118.4420	117.786	118.7654	116.0951
4	-	-	119.9303	114.7740
5	116.7785	114.308	114.2587	116.0417
6	117.4069	113.885	116.6304	114.9324
7	-	-	116.3753	116.613
8	114.6833	114.110	114.4472	115.2210
9	115.0344	116.648	115.8887	114.8816
10	-	-	116.5826	115.2499
11	115.8755	114.431	115.7893	115.017
12	115.2095	115.962	117.4369	115.9384
13	-	-	116.8799	116.1166
14	115.7217	115.154	115.0408	116.186
15	115.1445	115.276	116.8458	116.255

launch time by shifting to lower manning levels (until manning level 3), nor do you gain any performance by shifting the manning level to higher levels, as long as you are above three. To compare the experiments numerically, Table 4.7 shows the means for each of the experiments across manning levels.

The Pareto frontier shown in Figure 4.11 hugs the bottom of the scatter plot at the minimum mean values. From Table 4.7, the "optimal" choice corresponding to the highest performance, chosen the lowest average launch times, can be determined. The minimum values are listed in column 1 of Table 4.8. The argument, now, can be made that at any given manning level, if one were to choose the conditions corresponding to the experiment with the lowest mean value at that manning level, then it would be the optimal weighting between cost & performance. In order to accept this argument, though, it is necessary to show that these are statistically significant decision points and did not "by chance" turn out to be the optimal points. Figure 4.12 shows a pairwise significance comparison between all of the manning levels and conditions with Bonferroni correction for family wise error.

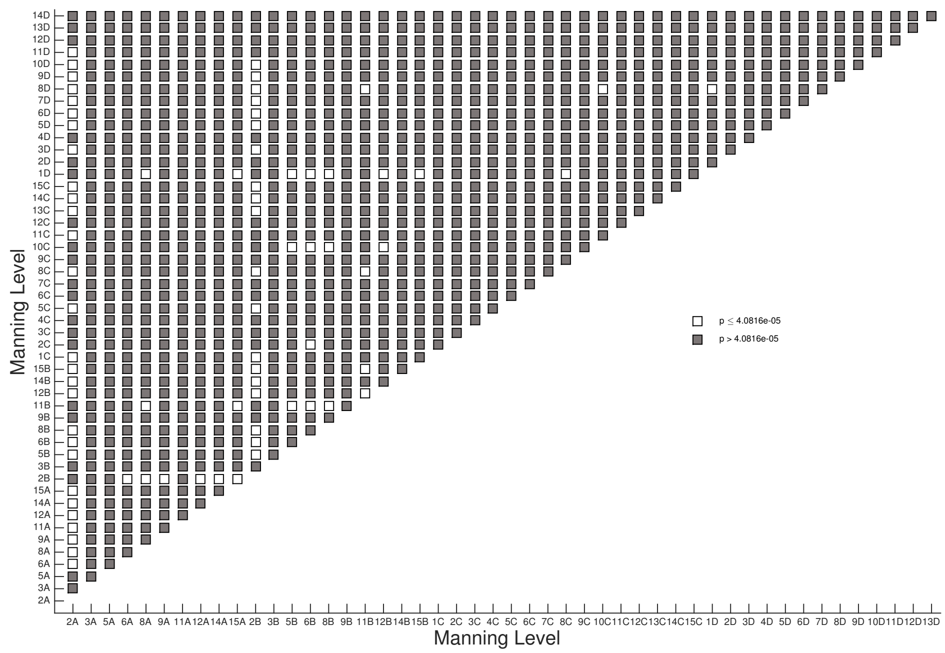

FIGURE 4.12: Pairwise Comparisions across Experiments

At each manning level, it would be necessary that the optimal choice be significantly different than all other choices at that manning level, to make an argument for using it. Table 4.8 shows the statistical significance from Figure 4.12 of the optimal manning levels against the other choices at that manning level.

Note that none of them are statistically significant from the other choices in the manning level. This is to say that, at any given manning level, it does not matter, based on mean total launch time, which experimental conditions are chosen. They are all statistically equivalent. This is surprising and suggests there is another bottleneck in the system. [46] has determined that the holdback bar installation, which takes a significant percentage of the catapult preparation time, may be a bottleneck and wash out any other inefficiencies in the aircraft preparations. For an analysis on how

Table 4.8: KHL Pairwise Test on Minimum Mean Launch Times

Level	A	B	C	D
1C	-	-	-	N
2D	N	N	N	-
3D	N	N	N	-
4D	-	-	N	-
5C	N	N	-	N
6B	N	-	N	N
7C	-	-	-	N
8B	N	-	N	N
9D	N	N	N	-
10D	-	-	N	-
11B	N	-	N	N
12A	-	N	N	N
13D	-	-	N	-
14C	N	N	-	N
15A	-	N	N	N

these optimal manning results would change if the hold back bar installation were hastened, see Appendix C.

This analysis concludes that there is little to no improvement in performance to be made across any of the experimental conditions. Costs may be minimized at no cost to performance by choosing manning level 3 for any of the conditions. In Chapter 5, overall conclusions are drawn and recommendations for optimal manning on the carrier deck are discussed. Future directions of this research and other applications are also presented.

5

Conclusions and Future Work

In this work, the Personnel Multi-Agent Safety and Control Simulator was developed and positioned with respect to the other performance models in the field of human factors engineering. In particular, it focused on the introduction of aircraft carrier maintenance crew human performance models and aircraft failure models in four phases of the launch cycle. The total launch time of a 22 aircraft launch cycle is used as a the performance metric to compare results across four experiments varying operator skill level and level of automation used in repairs. No statistical significance was found between experimental conditions, and all experiments showed manning level 3 is the operating point with minimal cost that still has no trade off in performance, so this point is considered to be the optimal manning level.

It is the conclusion of this work that the current carrier deck operations is far too over staffed from the perspective of balancing costs against performance, when mean total launch time is the performance metric. Maintenance crews, even under the current skill separation conditions, can be reduced to approximated two crews per 22 aircraft. This does not, however, take into concerns of preparedness for contingency operations during surges or deck failures. This is a question that can be addressed in future work, as discussed below. Furthermore, hastening Phase 4 failure repair times and training maintenance crews for multiple aircraft types does not improve performance unless an automated holdback bar is introduced, hastening the launch preparation cycle on the catapult, as discussed in Appendix C. The following sections discuss directions for future research on this work.

5.1 Future Work

PMASCS laid the framework for examining optimal manning on aircraft carrier decks. It is by no means exhaustive and may be extended and improved upon in many ways. The performance metric of mean total launch time is quite simple, and may not necessarily be the only metric of interest in determining flight deck performance. Also, this work focused mainly on one operator type. The human performance modeling presented herein can be built upon to address many other aspects of optimal manning on carrier decks and in other regimes. The following sections discuss the direction that I suggest future researchers consider with this project.

5.1.1 Improved Performance Models for non-Maintenace Crews

The extension of MASCS to PMASCS increased the number of simulated operator type by five. This now encompasses the majority of the crew types on an aircraft carrier related to launch operations of fixed wing aircraft. Some of the behaviors, especially of the ordnance crew and safety crew, have not been described in detail in this work as they do not pertain to the research questions. Their models could be extended to incorporate these operator behaviors and their impact on launch times. Furthermore, the analysis presented herein could be used to determine the optimal number of each type of operator type on deck, which may in turn identify other bottlenecks in the system.

The maintenance crew models developed for PMASCS do not take into account the individual operators or their specific training sets, which may be present in current operations. Each squadron, hypothetically, may have specific maintenance operators for each subsystem of the aircraft. Including resolution of maintenance skill sets at this level in the performance models may provide insight into how to better train the operators to improve overall deck performance. It may not, as proposed in this work, be possible for one person (or crew) to be generically trained on all repairs for all aircraft. A solution may exist in which one person is trained to perform maintenance on a specific subsystem type on multiple aircraft types.

5.1.2 Recovery of Aircraft

In each of these experiments all four catapults were utilized for launching in a single launch cycle. In reality, all of these aircraft must be recovered and prepped for the next launch cycle or during another launch cycle. This requires that catapults 3 and 4 be unused during the recovery of aircraft since they occupy the landing strip. It is likely that the optimal number of maintenance crews would change depending on the number of available catapults. Furthermore, the recovery of aircraft is a highly dangerous activity on the carrier deck and this should be taken into account in any safety metrics that are developed. Area separation of aircraft and operators is all the more important when the landing strip must not be occupied, so this can also be added to PMASCS.

5.1.3 Additional Performance Metrics

Mean total launch times for 22 aircraft is only one metric for performance of a carrier deck. Safety of operators as measured by halo violations or otherwise would be another metric that could be considered in the Pareto analysis. Also, as mentioned above, it is necessary to be prepared for contingency operations. If the deck is manned with very few operators, it may be the case that a disaster like a fire could not appropriately be handled. It is an open research question as to how to model these sorts of events and whether or not automation could be introduced to still allow for a reduction in manning on the carrier deck. A performance function, analogous to a cost function in optimal control, could be conditioned on multiple variables such as these and this could lead to a very different Pareto front. If this Pareto front is convex, then the optimal solution can be found by finding the minimum of the frontier.

5.1.4 Extension of Work to Other Fields

This work can be extended to many fields other than the carrier deck as well. Any environment in which the number of personnel has a direct affect in the quality of performance, such as hospitals or manufacturing plants, could apply this type of model and analysis to help drive optimal manning decisions.

6

Appendix A

Table 6.1: Multiple Compare Data for Experiment A

Sample A	Sample B	min	estimate	max	p-value
1	2	23.750	87.113	150.475	0.001
1	3	44.220	109.497	174.773	0.000
1	4	62.941	128.952	194.962	0.000
1	5	81.681	147.691	213.701	0.000
1	6	75.619	139.577	203.534	0.000
1	7	54.750	118.113	181.475	0.000
1	8	70.824	134.187	197.549	0.000
1	9	59.935	123.298	186.660	0.000
1	10	80.783	147.584	214.385	0.000
2	3	-42.309	22.384	87.078	0.985
2	4	-23.595	41.839	107.273	0.583
2	5	-4.856	60.578	126.012	0.098
2	6	-10.898	52.464	115.827	0.208
2	7	-31.762	31.000	93.762	0.866
2	8	-15.688	47.074	109.836	0.342
2	9	-26.577	36.185	98.947	0.720
2	10	-5.761	60.471	126.703	0.109
3	4	-47.834	19.455	86.743	0.996
3	5	-29.095	38.194	105.483	0.738
3	6	-35.196	30.080	95.356	0.908
3	7	-56.078	8.616	73.309	1.000

Table 6.1 – *Continued from previous page*

Sample A	Sample B	min	estimate	max	p-value
3	8	-40.004	24.690	89.383	0.971
3	9	-50.893	13.801	78.494	1.000
3	10	-29.978	38.087	106.152	0.754
4	5	-49.262	18.739	86.740	0.997
4	6	-55.385	10.625	76.636	1.000
4	7	-76.273	-10.839	54.595	1.000
4	8	-60.199	5.235	70.669	1.000
4	9	-71.088	-5.654	59.780	1.000
4	10	-50.137	18.632	87.402	0.998
5	6	-74.124	-8.114	57.896	1.000
5	7	-95.012	-29.578	35.856	0.918
5	8	-78.938	-13.504	51.930	1.000
5	9	-89.827	-24.393	41.041	0.976
5	10	-68.876	-0.107	68.662	1.000
6	7	-84.827	-21.464	41.898	0.987
6	8	-68.753	-5.390	57.972	1.000
6	9	-79.642	-16.279	47.083	0.998
6	10	-58.794	8.007	74.808	1.000
7	8	-46.688	16.074	78.836	0.998
7	9	-57.577	5.185	67.947	1.000
7	10	-36.761	29.471	95.703	0.925
8	9	-73.651	-10.889	51.873	1.000
8	10	-52.835	13.397	79.629	1.000
9	10	-41.946	24.286	90.518	0.978

Table 6.2: Multiple Compare Data for For Experiment B

Sample A	Sample B	min	estimate	max	p-value
1	2	14.069	79.583	145.098	0.005
1	3	79.396	143.638	207.880	0.000
1	4	81.742	145.984	210.226	0.000
1	5	71.756	136.612	201.468	0.000
1	6	34.027	99.542	165.056	0.000
1	7	67.476	132.332	197.188	0.000
1	8	46.436	111.292	176.148	0.000
1	9	59.827	123.495	187.164	0.000
1	10	49.328	116.314	183.301	0.000

Table 6.2 – *Continued from previous page*

Sample A	Sample B	min	estimate	max	p-value
2	3	-0.188	64.054	128.297	0.051
2	4	2.158	66.401	130.643	0.036
2	5	-7.828	57.028	121.884	0.143
2	6	-45.556	19.958	85.473	0.994
2	7	-12.108	52.748	117.604	0.230
2	8	-33.148	31.708	96.564	0.873
2	9	-19.757	43.912	107.581	0.469
2	10	-30.256	36.731	103.718	0.776
3	4	-60.598	2.346	65.290	1.000
3	5	-70.597	-7.026	56.544	1.000
3	6	-108.338	-44.096	20.146	0.477
3	7	-74.877	-11.306	52.264	1.000
3	8	-95.917	-32.346	31.224	0.844
3	9	-82.501	-20.142	42.216	0.991
3	10	-93.067	-27.323	38.420	0.951
4	5	-72.943	-9.372	54.198	1.000
4	6	-110.685	-46.442	17.800	0.397
4	7	-77.223	-13.652	49.918	1.000
4	8	-98.263	-34.692	28.878	0.781
4	9	-84.847	-22.489	39.870	0.980
4	10	-95.413	-29.670	36.074	0.919
5	6	-101.926	-37.070	27.786	0.730
5	7	-68.471	-4.280	59.911	1.000
5	8	-89.511	-25.320	38.871	0.965
5	9	-76.107	-13.116	49.875	1.000
5	10	-86.640	-20.297	46.046	0.994
6	7	-32.066	32.790	97.646	0.849
6	8	-53.106	11.750	76.606	1.000
6	9	-39.715	23.954	87.622	0.974
6	10	-50.214	16.773	83.760	0.999
7	8	-85.231	-21.040	43.151	0.990
7	9	-71.827	-8.836	54.155	1.000
7	10	-82.360	-16.017	50.326	0.999
8	9	-50.787	12.204	75.195	1.000
8	10	-61.320	5.023	71.366	1.000
9	10	-72.364	-7.181	58.002	1.000

Table 6.3: Multiple Compare Data for Experiment C

Sample A	Sample B	min	estimate	max	p-value
1	2	-82.166	33.538	149.243	1.000
1	3	3.523	121.417	239.310	0.036
1	4	-32.685	85.208	203.102	0.476
1	5	76.348	190.143	303.937	0.000
1	6	22.219	137.923	253.627	0.005
1	7	34.684	136.269	237.854	0.001
1	8	76.933	181.227	285.522	0.000
1	9	34.464	138.000	241.536	0.001
1	10	34.487	146.600	258.713	0.001
1	11	46.223	147.808	249.393	0.000
1	12	14.789	117.625	220.461	0.009
1	13	67.573	171.109	274.645	0.000
1	14	67.213	169.400	271.587	0.000
1	15	64.748	164.759	264.769	0.000
2	3	-25.185	87.878	200.941	0.346
2	4	-61.393	51.670	164.733	0.969
2	5	47.822	156.604	265.387	0.000
2	6	-6.394	104.385	215.163	0.090
2	7	6.794	102.731	198.668	0.022
2	8	48.887	147.689	246.490	0.000
2	9	6.461	104.462	202.462	0.024
2	10	6.039	113.062	220.084	0.027
2	11	18.332	114.269	210.206	0.005
2	12	-13.174	84.087	181.347	0.182
2	13	39.570	137.570	235.571	0.000
2	14	39.287	135.862	232.436	0.000
2	15	36.952	131.220	225.488	0.000
3	4	-151.510	-36.208	79.093	0.999
3	5	-42.381	68.726	179.834	0.737
3	6	-96.556	16.506	129.569	1.000
3	7	-83.713	14.853	113.418	1.000
3	8	-41.545	59.811	161.167	0.798
3	9	-83.992	16.583	117.159	1.000
3	10	-84.202	25.183	134.568	1.000
3	11	-72.175	26.391	124.957	1.000
3	12	-103.646	-3.792	96.063	1.000
3	13	-50.883	49.692	150.267	0.941
3	14	-51.203	47.983	147.170	0.950
3	15	-53.601	43.342	140.284	0.974

Table 6.3 – *Continued from previous page*

Sample A	Sample B	min	estimate	max	p-value
4	5	-6.173	104.935	216.042	0.088
4	6	-60.348	52.715	165.777	0.963
4	7	-47.505	51.061	149.627	0.915
4	8	-5.337	96.019	197.375	0.086
4	9	-47.784	52.792	153.367	0.906
4	10	-47.993	61.392	170.777	0.852
4	11	-35.966	62.599	161.165	0.698
4	12	-67.438	32.417	132.271	0.999
4	13	-14.675	85.900	186.476	0.198
4	14	-14.995	84.192	183.378	0.206
4	15	-17.392	79.550	176.493	0.256
5	6	-161.002	-52.220	56.562	0.953
5	7	-147.498	-53.874	39.751	0.827
5	8	-105.473	-8.916	87.642	1.000
5	9	-147.881	-52.143	43.595	0.879
5	10	-148.497	-43.543	61.412	0.987
5	11	-135.960	-42.335	51.290	0.971
5	12	-167.498	-72.518	22.463	0.377
5	13	-114.772	-19.034	76.704	1.000
5	14	-115.021	-20.743	73.535	1.000
5	15	-117.299	-25.384	66.530	1.000
6	7	-97.591	-1.654	94.283	1.000
6	8	-55.497	43.304	142.106	0.978
6	9	-97.923	0.077	98.077	1.000
6	10	-98.345	8.677	115.699	1.000
6	11	-86.052	9.885	105.822	1.000
6	12	-117.558	-20.298	76.962	1.000
6	13	-64.815	33.186	131.186	0.998
6	14	-65.097	31.477	128.051	0.999
6	15	-67.433	26.836	121.104	1.000
7	8	-36.857	44.958	126.773	0.871
7	9	-79.115	1.731	82.577	1.000
7	10	-81.243	10.331	101.905	1.000
7	11	-66.794	11.538	89.871	1.000
7	12	-98.592	-18.644	61.303	1.000
7	13	-46.007	34.839	115.686	0.981
7	14	-45.981	33.131	112.242	0.986
7	15	-47.790	28.489	104.769	0.995
8	9	-127.453	-43.227	40.998	0.921

Table 6.3 – *Continued from previous page*

Sample A	Sample B	min	estimate	max	p-value
8	10	-129.198	-34.627	59.943	0.996
8	11	-115.235	-33.420	48.396	0.989
8	12	-146.965	-63.602	19.761	0.378
8	13	-94.344	-10.119	74.107	1.000
8	14	-94.389	-11.827	70.735	1.000
8	15	-96.321	-16.469	63.383	1.000
9	10	-85.133	8.600	102.333	1.000
9	11	-71.038	9.808	90.654	1.000
9	12	-102.787	-20.375	62.037	1.000
9	13	-50.176	33.109	116.393	0.991
9	14	-50.202	31.400	113.002	0.994
9	15	-52.100	26.759	105.617	0.998
10	11	-90.366	1.208	92.782	1.000
10	12	-121.934	-28.975	63.984	0.999
10	13	-69.225	24.509	118.242	1.000
10	14	-69.441	22.800	115.041	1.000
10	15	-71.666	18.159	107.983	1.000
11	12	-110.130	-30.183	49.765	0.995
11	13	-57.545	23.301	104.147	1.000
11	14	-57.519	21.592	100.704	1.000
11	15	-59.329	16.951	93.230	1.000
12	13	-28.928	53.484	135.896	0.664
12	14	-28.936	51.775	132.486	0.683
12	15	-30.804	47.134	125.071	0.767
13	14	-83.310	-1.709	79.893	1.000
13	15	-85.209	-6.350	72.509	1.000
14	15	-81.721	-4.641	72.438	1.000

Table 6.4: Multiple Compare Data for Experiment D

Sample A	Sample B	min	estimate	max	p-value
1	2	-68.299	27.667	123.632	1.000
1	3	57.727	158.933	260.139	0.000
1	4	78.394	179.600	280.806	0.000
1	5	57.414	156.688	255.961	0.000
1	6	77.661	178.867	280.073	0.000
1	7	92.052	189.588	287.124	0.000

Table 6.4 – *Continued from previous page*

Sample A	Sample B	min	estimate	max	p-value
1	8	83.145	179.111	275.077	0.000
1	9	106.132	193.423	280.715	0.000
1	10	80.362	167.654	254.945	0.000
1	11	79.309	168.292	257.274	0.000
1	12	62.733	150.840	238.947	0.000
1	13	55.518	144.500	233.482	0.000
1	14	51.516	138.808	226.099	0.000
1	15	53.380	143.304	233.228	0.000
2	3	24.656	131.267	237.877	0.003
2	4	45.323	151.933	258.544	0.000
2	5	24.243	129.021	233.798	0.003
2	6	44.590	151.200	257.810	0.000
2	7	58.789	161.922	265.055	0.000
2	8	49.795	151.444	253.093	0.000
2	9	72.253	165.756	259.260	0.000
2	10	46.484	139.987	233.491	0.000
2	11	45.541	140.625	235.709	0.000
2	12	28.908	123.173	217.439	0.001
2	13	21.749	116.833	211.917	0.003
2	14	17.638	111.141	204.644	0.005
2	15	19.672	115.638	211.603	0.004
3	4	-90.684	20.667	132.018	1.000
3	5	-111.843	-2.246	107.351	1.000
3	6	-91.418	19.933	131.284	1.000
3	7	-77.371	30.655	138.681	1.000
3	8	-86.433	20.178	126.788	1.000
3	9	-64.385	34.490	133.364	0.998
3	10	-90.154	8.721	107.595	1.000
3	11	-91.012	9.358	109.729	1.000
3	12	-107.689	-8.093	91.502	1.000
3	13	-114.804	-14.433	85.937	1.000
3	14	-119.000	-20.126	78.749	1.000
3	15	-116.835	-15.629	85.577	1.000
4	5	-132.510	-22.912	86.685	1.000
4	6	-112.084	-0.733	110.618	1.000
4	7	-98.038	9.988	118.014	1.000
4	8	-107.099	-0.489	106.121	1.000
4	9	-85.051	13.823	112.697	1.000
4	10	-110.821	-11.946	86.928	1.000

Table 6.4 – *Continued from previous page*

Sample A	Sample B	min	estimate	max	p-value
4	11	-111.679	-11.308	89.062	1.000
4	12	-128.355	-28.760	70.835	1.000
4	13	-135.470	-35.100	65.270	0.998
4	14	-139.667	-40.792	58.082	0.988
4	15	-137.502	-36.296	64.910	0.997
5	6	-87.418	22.179	131.776	1.000
5	7	-73.317	32.901	139.118	0.999
5	8	-82.354	22.424	127.201	1.000
5	9	-60.160	36.736	133.631	0.995
5	10	-85.929	10.966	107.862	1.000
5	11	-86.817	11.604	110.025	1.000
5	12	-103.478	-5.847	91.783	1.000
5	13	-110.609	-12.188	86.234	1.000
5	14	-114.775	-17.880	79.015	1.000
5	15	-112.656	-13.383	85.890	1.000
6	7	-97.305	10.722	118.748	1.000
6	8	-106.366	0.244	106.855	1.000
6	9	-84.318	14.556	113.431	1.000
6	10	-110.087	-11.213	87.662	1.000
6	11	-110.945	-10.575	89.795	1.000
6	12	-127.622	-28.027	71.569	1.000
6	13	-134.737	-34.367	66.004	0.998
6	14	-138.933	-40.059	58.815	0.990
6	15	-136.768	-35.562	65.644	0.998
7	8	-113.610	-10.477	92.656	1.000
7	9	-91.280	3.835	98.949	1.000
7	10	-117.049	-21.934	73.180	1.000
7	11	-117.965	-21.297	75.372	1.000
7	12	-134.612	-38.748	57.115	0.990
7	13	-141.757	-45.088	51.580	0.963
7	14	-145.895	-50.781	44.334	0.894
7	15	-143.820	-46.284	51.252	0.957
8	9	-79.191	14.312	107.815	1.000
8	10	-104.961	-11.457	82.046	1.000
8	11	-105.903	-10.819	84.264	1.000
8	12	-122.536	-28.271	65.994	1.000
8	13	-129.695	-34.611	60.473	0.996
8	14	-133.807	-40.303	53.200	0.981
8	15	-131.772	-35.807	60.159	0.995

Table 6.4 – *Continued from previous page*

Sample A	Sample B	min	estimate	max	p-value
9	10	-110.346	-25.769	58.808	0.999
9	11	-111.452	-25.131	61.190	1.000
9	12	-128.002	-42.583	42.836	0.936
9	13	-135.244	-48.923	37.398	0.842
9	14	-139.192	-54.615	29.962	0.672
9	15	-137.410	-50.119	37.173	0.829
10	11	-85.683	0.638	86.959	1.000
10	12	-102.232	-16.814	68.605	1.000
10	13	-109.475	-23.154	63.167	1.000
10	14	-113.423	-28.846	55.731	0.998
10	15	-111.641	-24.349	62.942	1.000
11	12	-104.597	-17.452	69.694	1.000
11	13	-111.822	-23.792	64.239	1.000
11	14	-115.805	-29.484	56.837	0.998
11	15	-113.970	-24.987	63.995	1.000
12	13	-93.486	-6.340	80.806	1.000
12	14	-97.451	-12.032	73.386	1.000
12	15	-95.643	-7.536	80.571	1.000
13	14	-92.013	-5.692	80.629	1.000
13	15	-90.178	-1.196	87.787	1.000
14	15	-82.795	4.497	91.788	1.000

Table 6.5: Multiple Compare Data for Auto Holdback Bar Installation, Mannning Level 3

Sample A	Sample B	min	estimate	max	p-value
1	2	6.590	31.851	57.112	0.0066
1	3	16.214	41.849	67.484	0.0002
1	4	4.299	29.221	54.142	0.0138
2	3	-15.806	9.999	35.803	0.7520
2	4	-27.727	-2.630	22.466	0.9932
3	4	-38.101	-12.629	12.844	0.5798

7

Appendix B

7.1 Kruskal-Wallis H Test

The Kruskal-Wallis H (KWH) test [28], also referred to as the one-way ANOVA on ranks test, is a non-parametric test used to show statistical significance between multiple groups. A non-parametric test must be used because the distributions of launch times are not normally distributed (which will be shown in each experiment). Four assumptions on the data must be met in order to justify using the KWH test over other statistical significance tests.

Assumption 1: *The dependent variable should be continuous or ordinal.* The dependent variable in this case is the total launch time, which is a continuous variable so this assumption is met.

Assumption 2: *The independent variable should consist of three or more independent groups.* The independent variable in this case is the number of maintenance crews, and in each set consists of 10 independent groups for Experiments A & B, and 15 independent groups Experiments C & D. So, this assumption is met.

Assumption 3: *Each group of observations must be independent from the other* *groups.* Since each of the manning levels is observed from independent simulation trials, with different settings, the observation sets are independent from one another and this assumption is met.

Assumption 4: *The distributions of each set must have the same "shape" or the same variability.* In each experiment, I show that the distributions are similar for many of the manning levels but differ for the lower manning levels. This means that the KWH test cannot be used to compare the medians of the data, but it is still fine to compare the mean ranks of the data, so this is what is done.

7.2 Data Distributions

The following histograms show the distributions of each of the experiments at every manning level. The left most on the top row is for manning level 1, and the right most on the bottom row is for manning level 15. Notice that the distribution "shape" is not the same for every manning level, so by Assumption 4 of the KWH test, we can only compare mean ranks.

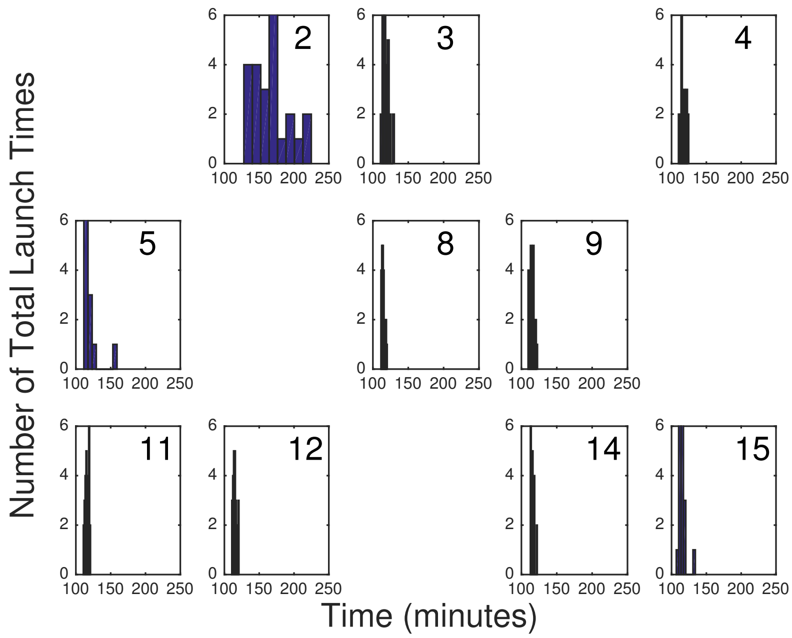

FIGURE 7.1: Launch Time Distributions for ExperimentA

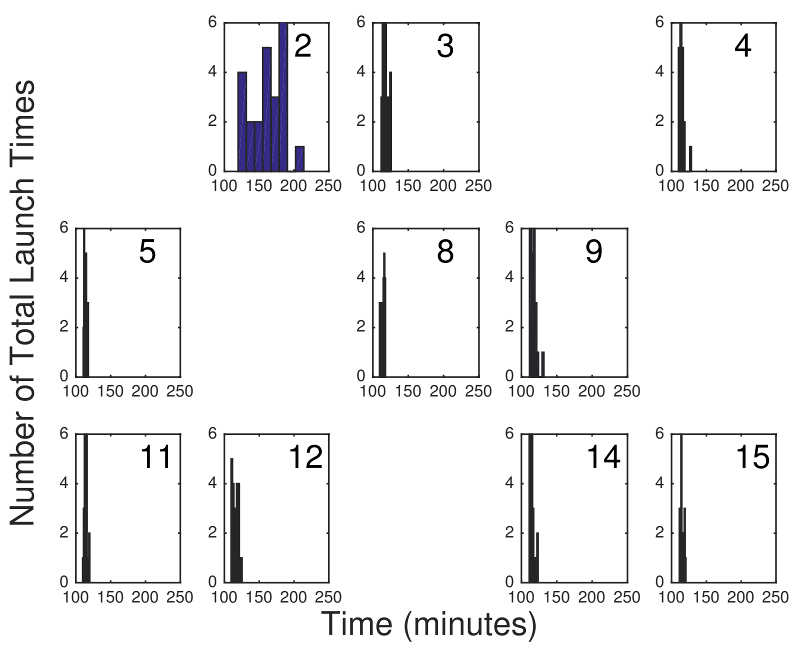

FIGURE 7.2: Launch Time Distributions for Experiment B

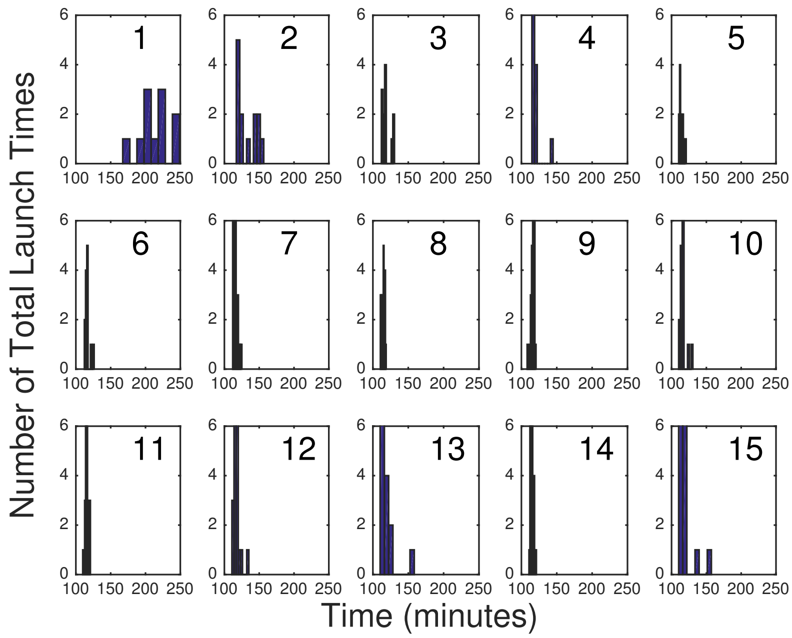

FIGURE 7.3: Launch Time Distributions for Experiment C

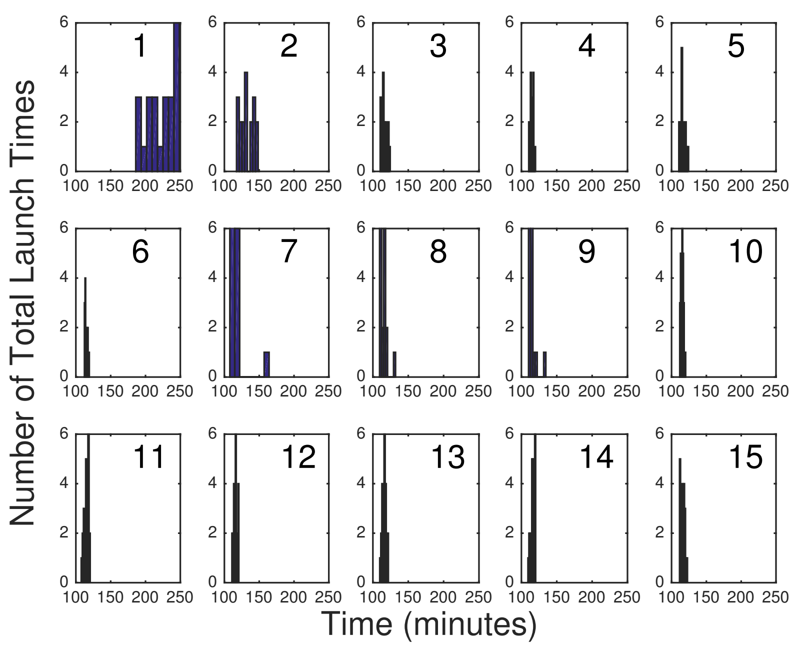

FIGURE 7.4: Launch Time Distributions for Experiment D

8

Appendix C: Automated Holdback Bar Pareto Analysis

The results in Chapter 4 show no change in mean total launch times until very low manning levels are reached (manning level 3 or lower), and show that neither Experiments B, C, or D yielded results statistically significantly different than the nominal case of Experiment A. This suggests that there is a bottleneck in the launch cycle washing out any improvements gained from additional maintenance crews. It is likely that the holdback bar installation time, which happens during the preparation of the aircraft for launch on the catapult, contributes to this bottleneck. In this section, I provide a second Pareto analysis across experiments for a scenario in which the holdback bar installation is hastened through automation.

8.1 Holdback Bar Installation

Upon arrival at the catapult, the aircraft is attached to the catapult with a holdback bar and a series of checks are performed to ensure proper functionality. To install the holdback bar, a crew member kneels beneath the aircraft and attaches a holdback bar between the aircraft's landing gear and the deck. This holdback bar helps to ensure that there is proper pressure in the steam catapult to launch the aircraft before acceleration begins. Once the necessary force is applied from the catapult to launch the aircraft, the holdback bar releases and allows the aircraft to be accelerated and launched. The time for installation of the holdback bar in the original simulations was drawn from a normal distribution with $\mu = 300s$ and $\sigma = 10s$. This is a longer than normal time for holdback bar installation which has been observed to be between 60 and 120 seconds. This installation time was chosen to be longer in order to capture the additional time required for aircraft safety checks that are performed at the catapult prior to launch. In this section, the installation time of the holdback bar is hasted by 10x the originals speed. This may be achievable, say, with additional automation in the installation of the holdback bar and other safety checks that are done on the aircraft at the catapult.

8.1.1 Pareto Analysis

Each of the four experiments (A, B, C, & D) across manning levels 1 to 15 were simulated again with the hastened holdback bar installation time. The total mean launch times, as expected, were lower across almost all manning levels. This is clear in the dashed line in Figure 8.1, which has flat region for manning levels greater than 3, similar to the original results to the manual holdback bar scenarios as shown by the solid black line.

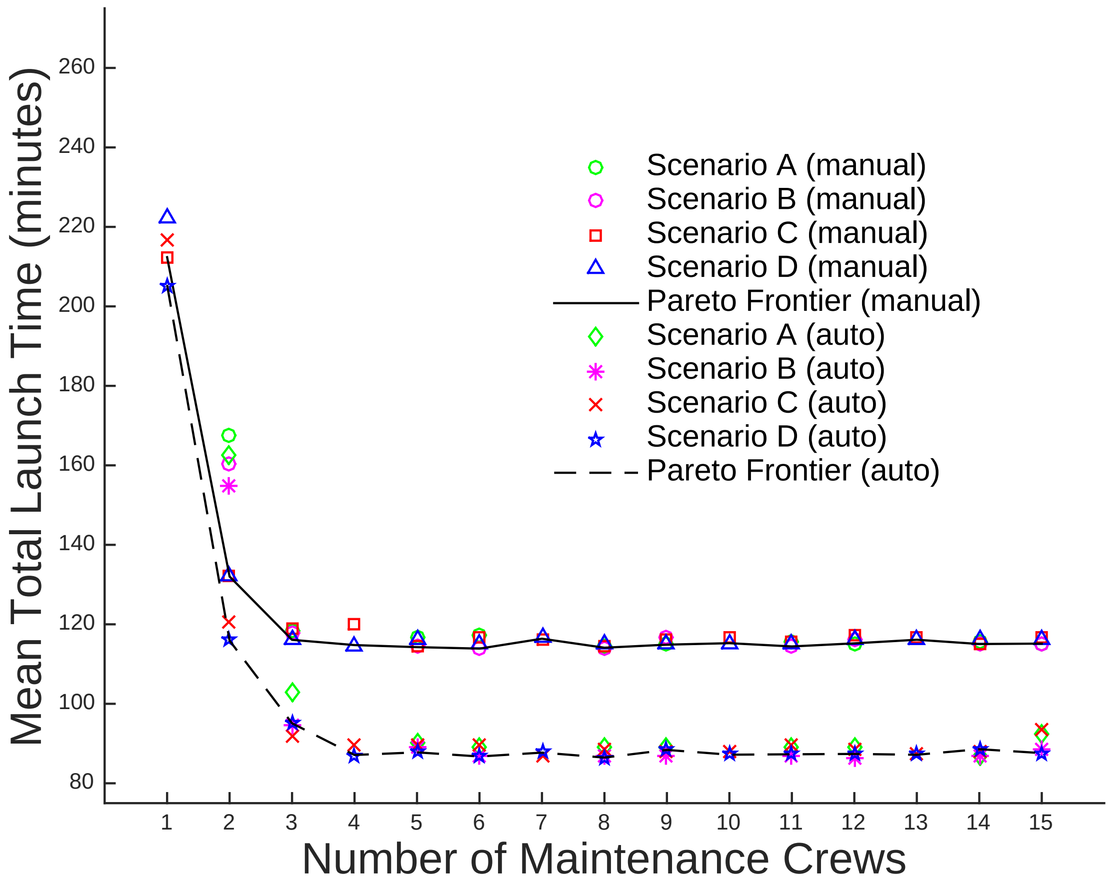

FIGURE 8.1: Pareto Frontier Analysis for an Automated Holdback Bar

The flat region for manning levels 3 to 15 is about 30 minutes lower for the automatic holdback bar than the manual holdback bar. This is expected since, on average, the holdback bar installation is 4.5 minutes faster for the auto case over the manual case and there are 22 aircraft being launched on 4 catapults in parallel. Equation 8.1 shows the expected time for auto vs manual holdback bar installation to be 24.75 minutes.

$$(300s - 30s) * (22 \text{ aircraft})/(4 \text{ catapults}) = 24.75 \text{ minutes} \tag{8.1}$$

The location of the upward trend elbow is between manning levels 3 and 4 for the auto holdback bar. Experiment A at manning level 3 has a notably higher mean total launch time than the other experiments' mean total launch times. The Kruskal Wallis H test shows that there is a statistical difference between the experimental conditions at manning level 3 with $p = 0.0002$.

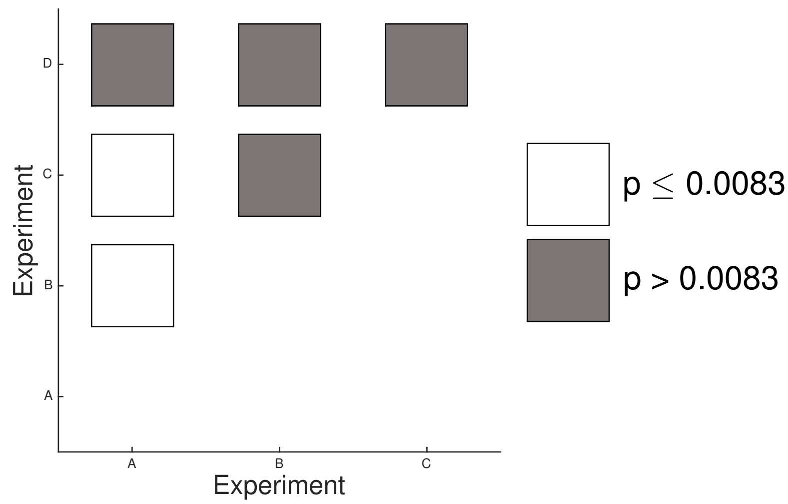

FIGURE 8.2: Pairwise Comparisons for Manning Level 3

Further analysis, as shown in Figure 8.2, shows that Experiment A is significantly different than experiments C and D, and marginally significantly different than Experiment D. The Bonferroni family wise error correction was used to determine the α level of $\alpha = 0.083$. The Bonferroni correction is a very strict approach to family wise correction, so it is noteworthy that the pairwise comparison between Experiments A and D had a significance level just above the α level. Full pairwise comparison statistics can be seen in Table 6.5 in Appendix A.

Bibliography

[1] Allan, R. J. (2009), "Survey of agent based modelling and simulation tools," Tech. rep., STFC Daresbury Laboratory.

[2] Allender, L. (2000), "Modeling human performance: Impacting system design, performance, and cost," *Simulation Series*, 32, 139–144.

[3] Bisseret, A. (1981), "Application of signal detection theory to decision making in supervisory control The effect of the operator's experience," *Ergonomics*, 24, 81–94.

[4] Bonabeau, E. (2002), "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences*, 99, 7280–7287.

[5] Boyer, M., Cummings, M. L., Spence, L. B., and Solovey, E. T. (2015), "Investigating Mental Workload Changes in a Long Duration Supervisory Control Task," *Interacting with Computers*, 27, 512–520.

[6] Byrne, M. D. and Pew, R. W. (2009), "A history and primer of human performance modeling," *Reviews of human factors and ergonomics*, 5, 225–263.

[7] Card, S. K., Newell, A., and Moran, T. P. (1983), *The psychology of human-computer interaction*, L. Erlbaum Associates Inc.

[8] Donmez, B., Nehme, C., and Cummings, M. L. (2010), "Modeling workload impact in multiple unmanned vehicle supervisory control," *IEEE Transactions on Systems, Man, and Cybernetics*, 40, 1180–1190.

[9] Doyle, M. R., Samuel, D. J., Conway, T., and Klimowski, R. R. (1995), "Electromagnetic aircraft launch system-EMALS," *Magnetics, IEEE Transactions on*, 31, 528–533.

[10] Edkins, G. D. and Pollock, C. M. (1997), "The influence of sustained attention on railway accidents," *Accident Analysis & Prevention*, 29, 533–539.

[11] Eubanks, J. L. and Killeen, P. R. (1983), "An application of signal detection theory to air combat training," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 25, 449–456.

[12] Fader, P. S. and McAlister, L. (1990), "An elimination by aspects model of consumer response to promotion calibrated on UPC scanner data," *Journal of Marketing Research*, pp. 322–332.

[13] Fong, T., Scholtz, J., Shah, J. A., Fluckiger, L., Kunz, C., Lees, D., Schreiner, J., Siegel, M., Hiatt, L. M., Nourbakhsh, I., et al. (2006), "A preliminary study of peer-to-peer human-robot interaction," in *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*, vol. 4, pp. 3198–3203, IEEE.

[14] Gilbert, N. and Terna, P. (2000), "How to build and use agent-based models in social science," *Mind & Society*, 1, 57–72.

[15] Hamilton, W. I. and Clarke, T. (2005), "Driver performance modelling and its practical application to railway safety," *Applied Ergonomics*, 36, 661–670.

[16] Hashemi, J., Spina, T. V., Tepper, M., Esler, A., Morellas, V., Papanikolopoulos, N., and Sapiro, G. (2012), "A computer vision approach for the assessment of autism-related behavioral markers," in *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on*, pp. 1–7, IEEE.

[17] Hayes, P. J. (1971), *The Frame Problem and Related Problems on Artificial Intelligence*, Stanford University, Computer Science Department.

[18] Hendrix, T. J. and Foreman, S. E. (2001), "Optimal long-term care nurse-staffing levels," *Nursing Economics*, 19, 164.

[19] Hennigan, W. (2013), "Navy Drone X-47B Lands on Carrier Deck in Historic First," *Los Angeles Times*.

[20] Hiltz, J. A. (2005), "Damage Control and Optimized Manning," Tech. rep., DRDC-Atlantic.

[21] Hood, L. (1998), "Agent-based modeling," in *Conference Proceedings: Greenhouse Beyond Kyoto, Issues, Opportunities and Challenges*, vol. 31.

[22] Hudlicka, E. (2007), "Reasons for emotions," *Integrated Models of Cognition Systems*, 1, 263.

[23] Jin Park, S., Kim, C.-B., and Chan Park, S. (1999), "Anthropometric and biomechanical characteristics on body segments of in a Large-Scale Naval Ship Design Program: Usability Testing of Complex Military SystemsDDG 1000," *Applied human science*, 18, 91–99.

[24] John, B. E. and Gray, W. D. (1995), "CPM-GOMS: an analysis method for tasks with parallel activities," in *Conference companion on Human factors in computing systems*, pp. 393–394, ACM.

[25] John, B. E. and Kieras, D. E. (1996), "Using GOMS for user interface design and evaluation: Which technique?" *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3, 287–319.

[26] Johnson, P. E. (2002), "Agent-Based Modeling What I Learned From the Artificial Stock Market," *Social Science Computer Review*, 20, 174–186.

[27] Kleinman, D., Baron, S., and Levison, W. (1970), "An optimal control model of human response part I: Theory and validation," *Automatica*, 6, 357–369.

[28] Kruskal, W. H. and Wallis, W. A. (1952), "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, 47, 583–621.

[29] LeBaron, B. (2002), "Building the Santa Fe artificial stock market," *Physica A*.

[30] Lynas, D. and Horberry, T. (2011), "Human factor issues with automated mining equipment," *Ergonomics Open Journal*, 4, 74–80.

[31] Macmillan, N. A. (2002), "Signal detection theory," *Stevens' handbook of experimental psychology*.

[32] Macy, M. W. and Willer, R. (2002), "From factors to actors: Computational sociology and agent-based modeling," *Annual review of sociology*, pp. 143–166.

[33] Majumdar, A. and Ochieng, W. (2002), "Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload," *Transportation Research Record: Journal of the Transportation Research Board*, pp. 58–69.

[34] Mattingley, J., Wang, Y., and Boyd, S. (2011), "Receding horizon control," *Control Systems, IEEE*, 31, 52–65.

[35] Mazzoleni, M. J., Battaglini, C. L., and Mann, B. P. (2015), "Modeling heart rate dynamics in response to changes in exercise intensity," in *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. V008T13A029–V008T13A029, American Society of Mechanical Engineers.

[36] Miettinen, K. (1999), *Nonlinear Multiobjective Optimization*, International Series in Operations Research & Management Science, Springer US.

[37] Naiem, A., Reda, M., El-Beltagy, M., and El-Khodary, I. (2010), "An agent based approach for modeling traffic flow," in *Informatics and Systems (INFOS), 2010 The 7th International Conference on*, pp. 1–6, IEEE.

[38] Nathman, J. (1999), "CV NATOPS MANUAL (NAVAIR 00-80T-105)," *Naval Air Technical Services Facility, Philadelphia, PA*.

[39] Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., and Zelevinsky, K. (2002), "Nurse-staffing levels and the quality of care in hospitals," *New England Journal of Medicine*, 346, 1715–1722.

[40] Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2008), "Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs," *Journal of Cognitive Engineering and Decision Making*, 2, 140–160.

[41] Pickup, S. L., Ferren, M., Grant, C., Harms, N., Krustapentus, J., LaCasse, M. J., and Landesman, J. (2011), "Military Readiness: Navy's Report to Congress on the Impact of Training and Crew Size on Surface Force Material Readiness," Tech. rep., DTIC Document.

[42] Quintana, V., Howells, R. A., and Hettinger, L. (2007), "User-Centered Design in a Large-Scale Naval Ship Design Program: Usability Testing of Complex Military SystemsDDG 1000," *Naval Engineers Journal*, 119, 25–33.

[43] Ranyard, R. H. (1976), "Elimination by aspects as a decision rule for risky choice," *Acta Psychologica*, 40, 299–310.

[44] Rashford, S. and Myers, C. (2004), "Optimal staffing of helicopter emergency medical services is controversial," *Emergency Medicine*, 16, 269–270.

[45] Rodney, D. M., Bowes, M. D., Duquette, C. M., Russell, S. M., Lawler, K. S., and Oi, J. S. (2009), "Impact of Manning and Infrastructure Initiatives on the Surface Navy," Tech. rep., Center for Naval Analysis.

[46] Ryan, J. C. (2014), "Evaluating Safety Protocols for Manned-Unmanned Environments through Agent-Based Simulation," Ph.D. thesis, Massachusetts Institute of Technology.

[47] Ryan, J. C. and Cummings, M. L. (2014a), "Development of an Agent-Based Model for Aircraft Carrier Flight Deck Operations," *M&S Journal*, pp. 5–15.

[48] Ryan, J. C. and Cummings, M. L. (2014b), "A Systems Analysis of the Introduction of Unmanned Aircraft Into Aircraft Carrier Operations," *Human-Machine Systems, IEEE Transactions on*, PP, 1–12.

[49] Scharf, L. L. (1991), *Statistical signal processing*, vol. 98, Addison-Wesley Reading, MA.

[50] Schmidt, D. K. (1978), "QUEUEING ANALYSIS OF THE AIR TRAFFIC CONTROLLER'S WORK LOAD," *IEEE Transactions on Systems, Man, and Cybernetics*, 8, 492–498.

[51] Schneider, B. (2008), "The Reference Model SimPan - Agent-Based Modelling of Human Behaviour in Panic Situations," in *Computer Modeling and Simulation, 2008. UKSIM 2008. Tenth International Conference on*, pp. 599–604.

[52] Scott, C. and Nowak, R. (2005), "A Neyman-Pearson approach to statistical learning," *Information Theory, IEEE Transactions on*, 51, 3806–3819.

[53] Seow, S. C. (2005), "Information theoretic models of HCI: a comparison of the Hick-Hyman law and Fitts' law," *Human-Computer Interaction*, 20, 315–352.

[54] Shemwell, D. M. (1994), "Method and apparatus for aiding a landing aircraft," US Patent 5,287,104.

[55] Sivalingam, R., Cherian, A., Fasching, J., Walczak, N., Bird, N., Morellas, V., Murphy, B., Cullen, K., Lim, K., Sapiro, G., et al. (2012), "A multi-sensor visual tracking system for behavior monitoring of at-risk children," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pp. 1345–1350, IEEE.

[56] Spindel, R. C., Laska, S., Cannon-Bowers, J. A., Cooper, D. L., Hegmann, K. C., Hogan, R. J., Hubbard, J. E., Johnson, J. A., Katz, D. J., Kohn Jr, R., et al. (2000), "Optimized Surface Ship Manning," Tech. rep., DTIC Document.

[57] Studios, G. T. (2016), *Golden T Game Engine GTGE: Game Programming for Java Programmer*.

[58] Swaminathan, J. M., Smith, S. F., and Sadeh, N. M. (1998), "Modeling supply chain dynamics: A multiagent approach*," *Decision sciences*, 29, 607–632.

[59] Thorpe, S., Fize, D., Marlot, C., et al. (1996), "Speed of processing in the human visual system," *nature*, 381, 520–522.

[60] Tversky, A. (1972), "Elimination by aspects: A theory of choice." *Psychological review*, 79, 281.

[61] Wallis, T. S. and Horswill, M. S. (2007), "Using fuzzy signal detection theory to determine why experienced and trained drivers respond faster than novices in a hazard perception test," *Accident Analysis & Prevention*, 39, 1177–1185.

[62] Wei, H., Ross, W., Varisco, S., Krief, P., and Ferrari, S. (2013), "Modeling of human driver behavior via receding horizon and artificial neural network controllers," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, pp. 6778–6785, IEEE.

[63] Young, W., Richardson, A., Ogden, K., and Rattray, A. (1982), "Road and rail freight mode choice: application of an elimination-by-aspects model," *Transportation research record*.