SYSTEM DESIGN CONSIDERATIONS FOR HUMAN-AUTOMATION FUNCTION ALLOCATION DURING LUNAR LANDING

A Thesis
Presented to
The Academic Faculty

by

Zarrin K. Chua

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology
August 2013

SYSTEM DESIGN CONSIDERATIONS FOR HUMAN-AUTOMATION FUNCTION ALLOCATION DURING LUNAR LANDING

Approved by:

Dr. Karen M. Feigh, Advisor
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Nicholas J. M. Patrick
Self-representing

Dr. Robert D. Braun
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Amy R. Pritchett
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Chirold Epp
Project Manager, ALHAT
NASA Johnson Space Center

Date Approved:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ΔV	Change in Velocity
a_d	Descent acceleration
D	Landing Site Diameter
$G_{x,y,z}$	Linear Acceleration in the x, y, z direction
m	Mass
P_d	Desired polynomial form on the divert trajectory
P_n	Desired polynomial form of the nominal trjactory
r	Current position
r_T	Targeted position
t	Time
t_{go}	Time to end of subphase
v	Current velocity
v_d	Desired descent rate
v_T	Targeted velocity
V_h	Horizontal Velocity
V_V	Vertical Velocity
ACT-R	Adaptive Control of Thought-Rational
AFM	Autonomous Flight Manager
ALHAT	Autonomous Landing and Hazard Avoidance Technology

ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ap	Apollo-like
b	Baseline
C	Computer
db	decibel
DEM	Digital Elevation Map
DOUG	Dynamic Onboard Ubiquitous Graphics
DSKY	Display and Keyboard unit
dV	Change in Velocity
E	East
EDGE	Engineering DOUG Graphics for Exploration
FLOAAT	Function-Specific Level of Autonomy and Automation Tool
GA	Genetic Algorithm
GBPC	Genetics-Based Policy Capturing
GNC	Guidance, Navigation, and Control
H	Human
HIDH	Human Integration Design Handbook
HITL	Human-in-the-Loop
HUD	Heads Up Display
IMU	Inertial Measurement Unit

ISC	Inertial Stellar Compass
LEM	Lunar Excursion Module
li	LIDAR warning
LIDAR	LIght Detection and RADAR
LMP	Lunar Module Pilot
LPD	Landing Point Designation
LPR	Landing Point Redesignation
LSS	Landing Site Score
MABA-MABA	Men Are Better At, Machines Are Better At
MANOVA	Multivariate Analysis of Variance
MIDAS	Man-Machine Integration Design and Analysis System
mo	Moderate
NASA	National Air and Space Administration
no	Nominal
OODA	Observe, Orient, Decide, Act
OV	Overview Display
P64	Landing Maneuver Approach Phase
P65	Velocity Nulling Guidance
P66	Rate of Descent
P67	Manual Guidance
PDI	Powered Descent Initiation

PFD	Primary Flight Display
POI	Point Of Interest
ROC	Reconfigurable Operational Cockpit
RPDM	Recognition-Primed Decision Model
RSE	Response Surface Equation
s	Shallow
W	West

SUMMARY

A desire to advance humanity's presence in space prompts the need for improved technology to send crew to places such as the Moon, Mars, and nearby asteroids. Landing at these locations will require vehicle capabilities greater than that previously used during the Apollo program or in Low Earth Orbit. Safely placing a crewed vehicle in any landing condition requires a design decision regarding the distribution of tasks between the crew and automation. A joint human-automation system is hypothesized to provide several advantages to an all-automation or all-human system by leveraging the computational power of automation with the creativity and flexibility of humans. However, knowing when and what to automate is a difficult question. A number of function allocation strategies have been suggested through the literature, but many are applicable to nominal, baseline operations only. Furthermore, these strategies provide little guidance for specific applications. As such, functions may be automated or assigned to the human without fully understanding the human's needs, limitations, and capabilities, or the impact of the allocation on mission performance.

In this thesis, the use of a cognitive process model is proposed for determining the necessary automated functionality to support astronaut decision making. A process model lacks the detailed modeling applied in cognitive architectures, focusing instead on the composite behavior and interactions of individual cognitive processes. This model formulation informs design suggestions suitable for the requirements analysis design stage. To construct this model, however, the current literature lacks sufficient detailed knowledge regarding astronaut decision making during the landing point designation task, and observations of astronauts landing on the Moon are not readily available. Therefore, the most feasible option is conducting a human-in-the-loop (HITL) experiment to examine two representative human-automation function allocations. The data collected in the HITL study informs the cognitive process model and also the requirements analysis.

The research conducted in this thesis is comprised of three complimentary research endeavors. The first endeavor consisted of characterizing landing point designation (LPD) through three HITL experiments, one of which was conducted with the NASA Astronaut Office at Johnson Space Center. The data from this experiment was used to characterize the decision making strategies employed by the astronauts and observe the effects of automated assistance, trajectory, and scenario on landing performance. The second endeavor was focused on developing a computational, rule-based cognitive process model of astronaut decision making through a variety of function allocation, environmental, and mission operations parameters based on empirical data. The model was validated by comparing model and experiment results and on a set of one thousand randomly generated landing scenarios. The third endeavor was concerned with developing system requirements for an automated system to assist astronaut decision making during both nominal and off-nominal landing scenarios and the analysis of cockpit display usage and information needs. A normative model is suggested for future space exploration.

The thesis has provided the following contributions:

1. Characterize human-system decision making during landing point designation and provide a quantitative database of this performance.

2. Develop a cognitive process model to establish performance benchmarks and expected achievements during conceptual design

3. Provide system requirements regarding information needs and cockpit design requirements for humans and automated systems during landing point designation.

ACKNOWLEDGEMENTS

This thesis is the culmination of sweat, blood, and tears, love, patience, and happiness. I could not have done this alone and so many people have helped me along the way. However, there are a few I would especially like to thank. First, my advisor **Dr. Karen Feigh**, who has been more than a technical supervisor, but a great mentor as well. She has introduced me to many innovative ideas of conducting research, studying, and teaching, and I do not believe I could have seen these perspectives elsewhere. I have learned so much from her and I sincerely hope we continue working together in the future. Thank you for being a such a strong role model for me.

My committee members, who proved to be an invaluable resource of knowledge and advice. **Dr. Robert D. Braun**, who co-advised me during my time at Georgia Tech and allowed me to run free with my research ideas in my early graduate school years. **Dr. Amy Pritchett**, whom I have long admired and shaped my viewpoints in cognitive engineering - every time she speaks and writes, I learn something new. **Dr. Chirold D. Epp**, who is the epitome of a research grandfather, the guru with endless patience and kindness. And **Dr. Nicholas J. M. Patrick**, a man of worldly perspectives and continued advice.

My technical mentor at NASA Johnson Space Center, **Mr. Robert L. Hirsh**, who is the most generous, patient, and ubiquitous mentor ever. He has continuously made himself available to answer all my questions and concerns and a substantial amount of software and hardware development used in this thesis could not have been completed without him. **Brett Sommers** was an incredible resource and simulation engineer during the experiment trials, without his help and dedication I would have not finished up the study in time. Likewise, the members of the **Autonomous Landing and Hazard Avoidance Team (ALHAT)** have all been gracious enough to help me on this academic journey.

I have been fortunate enough to be part of two academic families, the **Cognitive Engineering Center** and the **Space Systems Design Laboratory**, the members of which

have provided me the camaraderie and the ideal work environment. Thank you to my fellow academic brothers and sisters and members of the Old Ladies Republic: **Alexandra Coso**, **Brad Steinfeldt**, **Chris Cordell**, **Scottie-Beth Fleming**, **So Young Kim**, **Michael Grant**, **Vlad Popescu**. A big thank you to my undergraduates who have contributed to this thesis: **James Tolbert**, **Vidhur Vohra**, **Jessica deRenzy**, **Philip Dewire**, and **Jesse Tapia**.

I could not have gone through these last six years without the emotional support from the following individuals: **Eliot Quon**, my roommate, high-five partner, and consolation buddy; **Ludovic Hivin**, **Ludivine Ventura**, **Owen Beste**, **Holly Jeffrey**, **Jennie Bartell**, **Aude Marzuoli**, **Amanda Chou**, **Kemp Kernstine**, **David Torello**, **Shaun, Ashley, & Faith Tarpley**, **Aaron Brogley**, **Anna Kryachkova Kores**, and **Tobie Hale Toy**. Merci á tous ceux de **"l'Ambassade Française"** (mon surnom pour la communauté des étudiants Français et **Dr. Eric Feron**), pour avoir partagé nos deux cultures. Lastly, thank you to **Mom and Dad**, who never stopped believing in me, **Adib** and **Navid**, providing fountains of support for their older sister.

I also thank the NASA Graduate Student Researchers Program (GSRP) and the Georgia Space Grant Consortium for their financial support of this thesis. Providing the resources and means to explore innovative technologies and cutting-edge science only allows our society to grow stronger and wiser. Thank you and may future students also benefit from the generosities of your programs!

For the thirty-two whose dreams were brutally cut short on April 16, 2007: This one's for you. You all hold a place in my heart and are forever missed by Hokie nation.

CHAPTER I

INTRODUCTION AND MOTIVATION

A desire to advance humanity's presence in space prompts the need for improved technology to send crew to places such as the Moon, Mars, and nearby asteroids. Landing at these locations will require vehicle capabilities greater than that previously used during the Apollo program or in Low Earth Orbit. In particular, automation needs to be capable of working in tandem with onboard astronauts in high-stakes, time-critical situations. Since the last crewed landing on another celestial body during the Apollo program, there have been substantial technological improvements, particularly in the area of automation. Sensor technology can now scan, process, and map landing terrain for navigational use, increasing vehicle state forecasting ability and improving knowledge of ownship state. Smart guidance algorithms are now placing vehicles within meters of the intended target. Improved display functionality is streamlining information and easing pilot workload. However, these automation systems may fail or work less robustly when operating beyond their design boundaries - a legitimate problem facing potential landing sites where the accuracy of information cannot be verified. The poles and the dark side of the Moon are dimly lit - resulting in poor contrast conditions - and are punctuated by large obstacles such as craters and rocks [1]. The Martian atmosphere is not dense enough to fully decelerate a vehicle but imposes enough friction on the vehicle that extensive heat shielding or dissipating designs are necessary for survival. Strong winds, craters, and rocks, challenge the vehicle's descent to the surface [2]. Lastly, low atmospheric densities and significant topographical features of asteroids threatens the safety of crew and vehicle alike. These challenges require careful attention to the allocation of functions between the crew and automation.

Safely placing a crewed vehicle in any landing condition requires a design decision regarding the distribution of functionality between humans and automation. A joint human-automation system is hypothesized to provide several advantages to an all-automation or

all-human system, by leveraging the computational power of automation with the creativity and flexibility of humans. However, knowing when and what to automate is a difficult question. A number of function allocation strategies have been suggested through the literature, but many are applicable to nominal, baseline operations only. Furthermore, these strategies provide little guidance for specific applications. As such, functions may be automated or assigned to the human without fully understanding the human's needs, limitations, and capabilities. For challenging tasks such as landing point designation (LPD), the event of deciding where to land the vehicle, poorly designed function allocations can result in failure or severe retardation in system performance. Therefore, the primary research question is posed: How should the functionality of automated landing systems and crew be assigned to support mission operations during landing point designation? Answering this question also demands understanding, how does landing point designation performance vary in each of these function allocations?

1.1 Using Cognitive Models Within Systems Design

These research questions can be answered in a number of ways, such as ethnographies, operations observations, and historical analysis. In this thesis, a cognitive process model is proposed to determine the necessary automated functionality to support astronaut decision making. A number of models could be used during this design process, ranging from primitive models (low-level cognitive phenomena) to process models (high-level cognitive processing functions) to cognitive architectures (precise modeling of both low- and high-level behaviors). Cognitive process models are the most appropriate for this thesis. The focus is on establishing design guidelines for requirements analysis, the earliest stage of design, with the goal of determining functions appropriate for humans and automation systems as to support mission operations.

This cognitive process model also provide predictive capability early in the design cycle. Understanding how performance varies with respect to changes in automation or mission operations allows for better forecasting ability and iterative design cycles. In the case of

this thesis, the cognitive process model allows for prediction of the distribution of potentially chosen landing sites at a specific place on the Moon. Developing this model, however, requires an improved understanding of function allocation during lunar landing point designation. Current literature lacks sufficient detailed knowledge regarding astronaut decision making during this task, and observations of astronauts landing on the Moon are not readily available. Therefore, this work conducts several human-in-the-loop (HITL) experiments to examine the impact of representative function allocations on system performance and selected landing site safety. The data collected in these HITL studies informs the cognitive process model and also the requirements analysis.

The inclusion of the cognitive process model within the design cycle differs from convention by accounting for human decision making in a realistic and sophisticated manner. Furthermore, the use of a cognitive process model allows mission designers to anticipate potential behavior in a variety of scenarios and develop additional training regimina to enhance mission performance. Inadequate designs of human and automation functionality has led to training humans to account for system deficiencies, along with training to achieve best performance. Specifically, this work examines astronaut decision making and variations in their cognitive processes due to different function allocations, landing trajectory, and mission operations. It accounts for the different decision making cues, or elements that are sensed from a display or through a window, the relative importance of such cues, and the different search methods used to find solutions. This thesis also examines cognitive processes that happens prior to and during operation, setting operator expectations for an anticipated future state or operation, and the incorporation of such expectations in the actual event.

The research contained in this thesis is composed of three complementary research endeavors. The first endeavor characterizes LPD by decomposing this main task into five fundamental subtasks. Each task can be allocated to either human or automation, allowing analysis of several function allocation combinations. Three allocation cases were chosen for further investigation in this thesis: *Apollo-like*, *moderate*, and *robotic*. Three HITL experiments were conducted to understanding the timing and interactions between human and

automation. The main experiment conducted for this thesis was completed in conjunction with the NASA Astronaut Office and evaluates the changes in astronaut decision making due to function allocation (moderate, Apollo-like), trajectory (baseline, shallow), and scenario (nominal, off-nominal). Both qualitative and quantitative data are collected, in terms of performance and astronaut decision making strategies.

The second endeavor developed and validated the cognitive process model. This rule-based computational cognitive process model is developed from literature, experimental data, observations, and debriefings with members of the Astronaut Office describing the task sequencing, the decision making cues, cue preferences, and search methods used by the astronauts. The cognitive functions reported by participants in the main experiment were codified and established in a rule-based algorithm. The model focuses on decision making behavior with appropriate assumptions for motor functions, memory, and task management. The model is validated by comparing participant site selection with cognitive model site selection based on the same decision making cue usage, preferences, and search method. The model was further validated by evaluating trends over 1000 simulated data points.

The third endeavor applies the cognitive process model to examine the information needs and cockpit design requirements to support the responsibilities of the human and the automation system at the Apollo-like, moderate, and robotic function allocations. Historical data, literature, experimental data, cognitive model development, and display evaluations from members of the Astronaut Office are used to inform the system design requirements and provide suggestions for training topics. System requirements for the automation functionality are proposed. Additionally, a normative model is prescribed for future automated landing systems that adjusts functionality of humans and automation between nominal and off-nominal scenarios. All of these analyses establish guidelines for requirements analysis and conceptual design.

1.2 Contributions

The research presented in this thesis is envisioned to provide three major contributions:

1. Characterize human-system decision making during landing point designation and

provide a quantitative database of this performance.

2. Develop a cognitive process model to establish performance benchmarks and expected achievements during conceptual design.

3. Provide system requirements regarding information needs and cockpit design requirements for humans and automated systems during landing point designation.

1.3 Research Overview

There are seven chapters in this thesis, as seen in Table 1. Chapter 1, this chapter, presents an introduction and poses the motivation for this body of work. Chapter 2, *Literature Review*, provides the necessary background to understand the methodology proposed in this research. Topics covered in this chapter include a discussion on the known limitations of crewed spaceflight; the definition and attributes of a cognitive architecture; the different methodologies of function allocation; and relevant studies. Chapter 3 discusses landing point designation and visualizing this task through task decomposition and task analysis. The major results of two HITL experiments related to this field are discussed, and the design of a third HITL experiment intended to inform the cognitive process model is presented. Chapter 4 discusses the results of the main HITL experiment conducted with the NASA Astronaut Office and the observed human-system decision making in the simulated lunar environment. Chapter 5 presents the development of the cognitive process model from experimental data and the validation. Chapter 6 highlights the information needs, display design suggestions, cockpit arrangements, desired function allocation, training topics, and derived system requirements for use in future crewed lunar landing. Lastly, Chapter 7 summarizes the entire thesis, with a discussion of future work.

Table 1: Thesis Organization

Ch. 1	Introduction and Motivation
Ch. 2	Literature Review
Ch. 3	Decision Making during Landing Point Designation
Ch. 4	Changes in Astronaut Decision Making
Ch. 5	Cognitive Process Model Development
Ch. 6	Suggested Automation Responsibilities and Crew Training Objectives
Ch. 7	Conclusions and Future Work

CHAPTER II

LITERATURE REVIEW

Determining automation system functionality for a crewed landing system requires an appreciation of several related topics. Fundamentally, this thesis touches on the topics of function allocation, cognitive modeling, and crewed spacecraft design. First, function allocation for any task can be performed using one of many available schema. Second, a number of modeling techniques have been designed to model human-automation interaction which can be used for a variety of purposes through the design cycle. Several of these techniques are validated and effective for design use. Third, decades of human spaceflight experience have established well-known standards and limitations for crewed spaceflight, especially for the safety of the onboard crew. In addition to these general design heuristics, there are a number of related studies in this area of human-automation function allocation during landing point designation (LPD). This chapter will review the major contributions within each of these particular fields.

2.1 Function Allocation

In this thesis, the term *function allocation* refers to the distribution of tasks between a human operator and an automated system. Other authors in literature have provided additional insights on function allocation. Feigh and Dorneich [3] categorize function allocation as task sharing and task offloading. Bailey [4] takes a broader perspective, and includes the analysis and description of system functions within this division of responsibilities between human and automation. Pritchett et al. [5] note that function allocation must account for the activities of both human and automation, rather than in broad terms such as levels of automation [6].

The task of function allocation is a critical step within any design process, particularly so during the design of human-automation systems. Generally, this task occurs once system objectives and performance specifications are set, and answers the question, "What

responsibilities should be performed by the user or the system?" [7]. Sequentially, function allocation is performed during the *conceptual design* stage, if one subscribes to the four-stage model of systems engineering: conceptual, preliminary, detailed, and final design. However, the activity itself does not occur once, but iteratively [4]. Pritchett et al. note that function allocation "is also often the only issue with human-automation interaction that *can* be addressed at the earliest design stages, that is before the interface and machine logic have been established" [5]. As discussed in Chapter 1, this thesis is focused on developing system requirements for function allocation and is intended for the early stages of design. This subsection discusses different allocation strategies and considerations for this design activity.

2.1.1 Allocation Strategies

Several methods have been proposed on how to divide work between humans and automation. These methods fall into four distinct families, with a fifth family representing the most current and advanced method of function allocation. These five families, sometimes referred to by different terms, are 1) comparison, 2) leftover, 3) economical, 4) humanized, and 5) flexible or *adaptive* allocation [4]. Comparison allocation is the assignment of functions based on known human and automation strengths. Leftover allocation prescribes an "automate what you can" approach, with remaining, or "leftover", tasks assigned to the human. In economical allocation, responsibilities are assigned as to minimize the overall system cost. Humanized allocation is the conjugate of leftover; humans are assigned meaningful tasks first, with remaining functions assigned to automation. Lastly, adaptive allocation provides a framework for the humans to allocate tasks in real-time, rather than assigning functions during system development. Adaptive allocation can be achieved by modifying the function allocation, the task scheduling, the interaction, or the content [3].

The concept of comparison function allocation was first introduced in the 1950s. Basic lists, such "Fitts' List" or MABA-MABA (Men Are Better At, Machines Are Better At), detailing human and automation strengths were proposed and applied to function allocation. Table 2 describes the MABA-MABA list proposed by Fitts. This list has evolved

over the years to reflect the maturation of computing capabilities. These comparison lists provide simplistic guidelines to partition work, but should be utilized as an initial approach to function allocation, rather than the sole basis. There are several shortcomings to comparison allocation. First, this practice does not provide comparable measures of interest beyond system performance (such as workload, human performance effects) [3]. Second, MABA-MABA lists assume that there are fixed strengths and weaknesses of humans and automation, thus, tasks can be easily substituted by humans or automation [8]. As Dekker and Woods writes, "Capitalising on some strength of automation does not replace a human weakness. It creates new human strengths and weaknesses - often in unanticipated ways" [9, 8]. As mentioned previously, the tasks assigned to automation result in transformed tasks for humans. Third, the human operator is ultimately responsible for the system, especially in failure. Comparison allocation assumes that humans are capable of taking over tasks assigned to automation (while maintaining their own responsibilities) at any point of operation.

Table 2: **Fitts' List: Men Are Better At, Machines Are Better At.** Adapted from Fitts [10].

Human	**Machines**
Detection	Speed
Perception	Power
Improvisation	Routinization
Long-Term Storage	Short-Term Storage
Induction	Deduction
Memory	Performance of Simultaneous Operations

The growing computing capabilities as recognized by the MABA-MABA list is reflected in the leftover allocation approach. This approach recommends that all functions that could be automated thus should be. The remaining functions are assigned to the human. As with comparison allocation, there are several shortcomings to relying on this single strategy. First, the human is assumed to be capable of completing the remaining tasks. Furthermore, the human is also assumed to want or have the motivation to do so [4]. Second, this allocation strategy neglects the interactions between different types of tasks such as implementation and planning. For example, in order to implement a course of action, one

must have involvement in the planning, or at minimum, be given the opportunity to review and comprehend the intended set of actions. Limiting or removing the human from these critical stages leaves unrealistic expectations during automation failure [3, 9].

A slight improvement over the leftover and comparison allocation strategies is the economic allocation, where tasks are assigned to achieve the minimum cost associated with selecting, execution, training, etc. While an intuitive approach, as most systems are constrained by a limited budget, this approach has its disadvantages. First, this approach, like the others, is too simplistic and 1-dimensional and can place the human operator in a precarious position, as the system is designed for minimizing costs, rather than an optimal or robust level of performance. Furthermore, just design costs are considered, instead of design and operational costs. The gains associated with shortening the life cycle may not counter those induced from operational or maintenance costs [4]. Second, economically allocating functions does not account for the workload imposed on the human or situation awareness requirements, leading to poor performance.

The humanized allocation strategy mirrors function allocation with the actual operation and assigns tasks to humans based on reasoning and logic. Each responsibility assigned to the human must be justified; that is, the task must have purpose to retain user motivation and leverage human strengths. Remaining functions are allocated to automation. Slight variations of humanized allocation are recommended for general practice. Chapanis argues that functions that must be allocated to humans or automation due to safety, human or technological limitations, or system requirements should be assigned first. The remaining functions should be analyzed and several different alternative human-automation configurations proposed. These configurations are then ranked based on scoring against different weighting criteria. The most effective configuration is based on the results of the weighting criteria and one that is most cost-effective [11].

2.1.2 Current Methodology

Three major function allocation methodologies exist in current practice today: the work proposed by Parasuraman et al.; Endsley and Kaber's levels of automation; and NASA's

FLOAAT utility. This three research exemplars and others often describe the allocation as levels of automation. However, *levels of automation* is misleading, as the relationship is non-linear [12, 13] and one-sided. That is, the focus is strictly on automation functionality whereas the human's role is also evolving. The uni-dimensional scale is too coarse sufficiently describe the distribution of work between both agents [5, 6]. Parasuraman et al. [12] introduced a second dimension, a four-stage model of human information processing. They propose that automation (and human capability) can be applied to four classes of functionality: information acquisition; information analysis; decision and action selection; and action implementation. Acquisition refers to sensing and registering data. Analysis is defined as inference and utilizing memory across data to establish predictions. Decision is the generation of action options, and action is the selection of the best choice amongst the options. Implementation is the execution of the chosen set of actions. As such, automation can be ranked as high or low across these four classes. Aspects of this work are based on Sheridan and Verplanck's levels of automation [14] (Table 3).

Table 3: **Levels of Automation of Decision and Action Selection.** Adapted from Parasuraman et al. [12].

	Role of Automation
10	The computer decides everything, acts autonomously, ignoring the human.
9	informs the human only if it, the computer, decides to
8	informs the human only if asked, or
7	executes automatically, then necessarily informs the human, and
6	allows the human a restricted time to veto before automatic execution, or
5	executes that suggestion if the human approves, or
4	suggests one alternative
3	narrows the selection down to a few, or
2	The computer offers a complete set of decision/action alternatives, or
1	The computer offers no assistance: human must take all decisions and actions.

Parasuraman et al. recommends a four-criterion set for primary evaluation of various combinations of human-automation allocation. These criteria are standard human performance evaluation metrics. *a*) Mental workload; *b*) Situation awareness; *c*) Complacency; *d*) Skill degredation. Full application of the model also requires the consideration of two other criteria, which composes of a secondary evaluation set: 1) Automation reliability; and

2) the cost of decision and action outcomes.

Endsley and Kaber [13] developed a ten-point automation scale that also allocates work across four-subfunctions: 1) Monitoring; 2) Generating; 3) Selecting; and 4) Implementing. This scale is different from Sheridan and Verplanck's scale as combined efforts of human and automation are considered, instead of just allocating functions to just one or the other. Table 4 lists these levels, including the allocations of work.

Table 4: **Endsley and Kaber's Ten-Point Levels of Automation.** The allocation of responsibilities is indicated by Human (H) or Computer (C) [13].

Level	**Monitor**	**Generate**	**Select**	**Implement**
1: Manual Control	H	H	H	H
2: Action Support	H/C	H	H	H/C
3: Batch Processing	H/C	H	H	C
4: Shared Control	H/C	H/C	H	H/C
5: Decision Support	H/C	H/C	H	C
6: Blended Decision Making	H/C	H/C	H/C	C
7: Rigid System	H/C	C	H	C
8: Automated Decision Making	H/C	H/C	C	C
9: Supervisory Control	H/C	C	C	C
10: Full Automation	C	C	C	C

Lastly, NASA has developed a Level of Autonomous Assessment Scale of eight combinations of human and automation control [15]. This scale are based on Boyd's OODA (Observe, Orient, Decide, Act) loop [16] (Table 5). When used in tandem with the Functional Level of Autonomy Assessment Tool (FLOAAT), a spacecraft system designer should be able to determine the appropriate level of autonomy. FLOAAT is a 35-question questionnaire that determines two level of autonomy limits: one for trust and the other a cost/benefit limit. The system designer's results to the questionnaire are mapped to the eight-point scale and a recommendation for function allocation is provided. While this approach conveniently describes function allocation in a quantitative manner, it is insufficient based on the arguments presented by Pritchett et al [5, 6]. Automation levels provides a one-sided account of the distribution of tasks to multiple entities and implies there is a natural linear progression of activity. Two systems at the same level of automation may be very different with regards to their functionality.

Table 5: **Level of Autonomy Assessment Scale.** H stands for Human; C for computer. Adapted from Proud et al. [15].

	Observe	**Orient**	**Decide**	**Act**
8	C gathers, filters, prioritizes data; not displayed to H	C predicts, interprets, integrates data; not displayed to H	C performs final ranking; not displayed to H	C executes automatically, no H input
7	C gathers, filters, prioritizes data; not displayed to H; "program functioning" flag displayed	C analyzes, predicts, interprets, integrates; displays to H only if within programmed context	C performs final ranking; displays reduced set of options to H without decision logic	C executes automatically, informs H if required; allows for emergency override
6	C gathers, filters, prioritizes data, displayed to H	C overlays predictions with analysis, interprets data; display all to H	C performs final ranking; displays decision logic to H	C executes automatically, informs H; allows for emergency override
5	C gathers info; displays non-prioritized, filtered info to H	C overlays predictions with analysis, interprets data. H shadows for contingencies	C performs final ranking, all and decision logic displayed to H	C allows H time to veto and emergency override
4	C gathers info; displays highlights non-prioritized, relevant info to H	C analyzes, predicts; H responsible for interpretation	H/C perform final ranking; C results are prime	C allows H time to veto and emergency override
3	C gathers info; displays unfiltered, unprioritized H; H prime monitor for info	C prime source of analysis, predictions; H responsible for data interpretation	H/C perform final ranking; H results are prime	C executes decision after H approval
2	H gathers and monitors info;	H prime source of analysis, predictions; H responsible for data interpretation	H performs final ranking; C used as tool for assistance	H prime source of execution
1	H only gathers and monitors info	H responsible for data analysis, predictions, interpretation	C doesn't assist or perform final ranking; H does all	H alone executes

2.1.3 Considerations for Function Allocation

Allocation of functions to automation and humans is not a simple, absolute division of work. This activity may require multiple iterations throughout the overall design process as the system evolves. There are several considerations that must be taken into account. Function allocation does not necessarily increase or decrease work for the human operator, but modifies the type and frequency of work [3, 9]. For example, if an automated system is tasked with gathering information regarding an unexpected landing terrain, the astronaut is not abstractly removed from the process, although he may not be performing the specific work itself. Instead, the astronaut is monitoring the data collected by the automation - he is still performing observable and quantifiable work, but it has been transformed from sensing and filtering to monitoring. Additionally, allocation of functions between different agents can results in *induced work*, or new tasks created due to the collaboration between different entities. Induced functions do not necessarily require extra resources such as workload or physical utilities, but are a remnant of agents having only partial knowledge of all states and aspects of the system [17].

The evolution of human operator work is one of many considerations regarding human-automation interaction. Others include the "responsibility-authority-double-bind", where the automation has sole authority, but the human is responsible for the automation's actions [18]; automation trust (too much reliance or lack of trust in its capabilities) [19]; or *brittle* automation (the automation's inability to operate reliably outside of its designed boundary conditions) [20]. Pritchett et al. [5] conducted a review of these concerns and have suggested five requirements when designing an effective function allocation:

1. Each agent must be allocated functions that it is capable of performing.
2. Each agent must be capable of performing its collective set of functions.
3. The function allocation must be realizable with reasonable teamwork.
4. The function allocation must support the dynamics of the work.
5. The function allocation should be the result of deliberate design decisions.

Regardless of the allocation strategy, overall mission performance due to specific function allocations need to be evaluated for effectiveness. To do so, the work, the agents, and the system must all be cohesively modeled and extensive analysis be applied [21]. Pritchett et al. developed eight metrics to evaluate the efficacy of a function allocation. These issues concern workload, stability of the work environment, mismatches of responsibility and authority, incoherency, interruptive automation, automation boundary conditions, function allocations limiting human adaptation to context, and mission performance [5]. Quantitative measures of function allocation such as these allow designers to understand the capability of both human and automation roles, further improving overall design of functionality in terms understood by system designers.

2.2 Cognitive Modeling

Modeling cognition is difficult. Unlike other activities or behaviors, cognition is not directly measurable, each individual is different, and people demonstrate inconsistent behavior. In an attempt to characterize cognition, different types of models have been developed for specific levels of analysis. Cognitive models can be grouped into three broad categories: primitive models, process models, and cognitive architectures. Primitive models are focused on describing the science behind specific cognitive phenomena (e.g., memory, perception). Process models explain the interaction between these specific cognitive phenomena and the resulting actions, but are less focused on the science (e.g., decision making, judgment). Cognitive architectures are composed of both primitive and process models, designed to accurately portray all aspects of the human body from low- to high-level cognitive and behavioral functions. Gray et al. describes three major differences between cognitive architectures and traditional cognitive models. First, cognitive architectures focus on integration, rather than "building microtheories of isolated phenomena or mechanisms." Second, in this shift to integration, designers have subsequently emphasized more on cognitive control, rather than just these phenomena. Lastly, cognitive architectures were intended for computational use, thus developing a virtual human in a simulated setting. The powerful capability of cognitive architectures allows designers to to simulate human performance, supporting design

activities such as rapid prototyping and supplementing evaluation or human-in-the-loop studies [22]. Sun points out that developing cognitive architectures forces logical, detailed explanations of phenomena rather than vague, theoretical concepts [23]. Furthermore, researchers must develop general theories, rather than specific, limited-application concepts [23].

Depending on the design cycle, certain models are more appropriate than others. Assuming a generic, four step systems design process consisting of 1) requirements analysis, 2) concept generation, 3) preliminary design, and 4) detailed design, the complexity of the cognitive model grows as the system design proceeds. While a cognitive architecture can be used during any stage of design, the rigor required for model development reduces its efficacy in requirements analysis. Similarly, primitive and process models are effective during the detailed design stage but may lack the modeling accuracy necessary for design refinement during the detailed design phase.

2.2.1 Primitive Models

Primitive models are derived from a wide variety of fields and may explain cognitive phenomena from the level of synapses firing and passing neurons to the psychological descriptions of how the brain processes information. These models describe both cognitive and physiological behavior. Several prominent examples from psychology and computer science explain primitive modeling. The work of Card et al. [24] with the Model Human Processor relates the functionality of the human brain to a computer. Timing estimations for fundamental, low-level actions have been solicited from experimental data. Fitts' Law provides a for hand pointing movement that can be extended for human-computer interaction. Based on the physical dimensions of the activity, one can determine a timing approximation for when this movement will be completed. On the psychological end, Miller's law notes that humans can store about seven pieces of information (with a standard deviation of two) in working memory [25]. The Gestalt principles explain how humans perceive visual elements and synthesize such elements [26]. These methods are a selection of many specific theories of cognitive and behavioral phenomena.

2.2.2 Process Models

Process models are cognitive models that focus on the overall process, less on the characterization of individual cognitive phenomena. They overlap with primitive models, specifically, in the discussion of certain information processing elements such as judgment and decision making. As with primitive models, multiple methodologies exist in literature. An example of process models includes the Man-Machine Integration Design and Analysis System (MIDAS). MIDAS, developed at NASA Ames Research Center, supports rapid-prototyping of human performance and assists in the design and evaluation of new operational systems and environments [27]. MIDAS is composed of several models: input, cognitive, output, and processing. The input model includes a model of the information flow that describes the elements exchanged between human and machine. The perception of information is explained by the cognitive model, which includes the perception and memory models. The perception model contains theories on visual and auditory perception. Once this information is acquired, it is filed in one of MIDAS' three types of memory: short-term, long-term, and long-term working. As the human interacts within the direct and indirect environment - described by the domain model - MIDAS' output model produces displays of the task network, anthropology metrics, and the overall mission performance. Designers use this stream of data to assess the efficacy of proposed designs and set expectations for nominal human performance. Lastly, MIDAS uses a processing model to manage simulated scenarios. The processing model is comprised of a task manager model, a model state, and a library of primitive models. The task manager model schedules and monitors human activities - continuously updating the model state. The primitive models are effectively fundamental models describing basic functions [28]. MIDAS is written in a combination of C/C++ and Lisp [27].

Another well-known model that could be considered a process model is the Lens Model [29] relates a known set of environmental cues and human judgment using multiple linear regressions. Correlations between the environment cues and the judgment provide commentary on the judgment quality, the knowledge of the judge, and the predictability of the environment. The Lens Model provides the mechanism to mathematically model how the

human uses cues in the environment and in his or her judgments. Recently, Rothrock and Kirlik have revisited Hammond's work and created a non-compensatory equivalent to the Lens Model, a method named Genetics-Based Policy Capturing (GBPC) [30]. GBPC uses a genetic algorithm (GA) to determine non-compensatory heuristics used by a judge in a defined scenario. GBPC postulates on a wide set of heuristic candidates and uses the GA to determine the one most likely to match the behavior of the judge.

Lastly, one such process-model particularly relevant to this work is the Recognition-Primed Decision Model (RPDM), which focuses specifically on the decision making behavior of experts. Developed by Klein [31], RPDM notes that experienced decision makers tend to generate, analyze, and select or reject possible decision options serially instead of in parallel. They select an acceptable rather than optimal decision option. The theory derives its name because the chosen decision option is generally derived from similar experiences recalled and recognized by the decision maker. RPDM is characterized by mentally simulating the decision option outcome, diagnosing potential shortcomings, and modifying the course of action as necessary.

2.2.3 Cognitive Architectures

Since the 1980s, cognitive architectures have been developed and matured or discarded. Gray, in his 2008 study, suggests there are at least 50 different architectures available [32]. However, there two that have developed as prominent, well-recognized mainstays: Soar and Adaptive Control of Thought-Rational (ACT-R). Soar and ACT-R are rule-based cognitive architectures that determine a set of actions apropos to human thinking based on user-established goals [33].

Soar, developed by the University of Michigan, is a production-based, deterministic system that models problem-solving. Driven by Principle P9 of the Model Human Processor [34], Soar assumes that humans apply knowledge in a rational manner and rely on symbolic information [33]. The problem, or task, is composed in a problem space where a large number of states exists. States contain descriptors of an aspect of the problem space. Soar models problem-solving by beginning with knowledge of an initial state and a goal state

and uses operators to move between states. Decision cycles are used to select operators and contains five phases: input, elaboration, decision, application, and output. Within this cycle, which last about 50 ms [35], Soar takes input from changes in perception to update the current state. Next, current knowledge contained within working memory is compared against procedural memory. All rules that match are fired in parallel, or implemented to change working memory. Included in these rules are preferences for which operator to select. The elaboration phase is complete when all matches have been fired. Soar then determines which operator to use, based on the preferences set by the elaboration phase. After selecting the appropriate operator, the current state shifts to a new space within the problem space, where an output, or feedback, is received. The decision cycle continues until the goal state is obtained [36].

On occasion, the architecture may reach an impasse, or when the decision cycle cannot select a new operator - there may be equal preferences or a lack thereof. During an impasse, Soar creates a substate with a new subgoal of selecting between operators. Instead of comparing working memory to procedural memory, Soar uses knowledge to compare different operators. After an operator is selected, a new rule is created in order to avoid this particular situation in the future. The subgoal is achieved and Soar continues the rest of the decision cycle. Essentially, Soar learns during impasses and improves its knowledge deficiency. The framework of Soar allows for flexibility, particularly for dynamic goals in complex environments. The learning during impasses makes Soar unique compared to other cognitive architectures. Soar is written in Python and Tool Command Language/Tk.

ACT-R, developed by Carnegie Mellon University, is another production-based, but probabilistic system that models problem-solving. The foundations of ACT-R is in psychological theory, with the functionality primarily based on experimental data. ACT-R produces quantitative measures such as time to perform a task and accuracy of task in order to compare against true human performance data. ACT-R consists of two modules: a perception-motor module contain visual and manual components; and a memory module that consists of declarative and procedural elements. Similar to Soar, ACT-R consists of production firing and matching. First, ACT-R searches for a rule, or a production, that

matches elements within declarative memory. A production rule is then selected based on the calculation of its expected utility. Lastly, the production rule is fired, and the appropriate actions are carried out in progressing towards the intended goal. One production rule is fired per cycle, thus, actions are completed serially. If more than one production rule matches the declarative memory, then conflict resolution must be carried out. The rules are assigned an expected utility based on three components: the probability that the goal is achieved if the rule is fired; the expected utility of the goal; and the cost of achieving the goal. The rule with the highest utility is selected [33] [35] [37]. ACT-R was originally developed in Lisp, but contains variants in Java and Python [37].

2.2.4 Validation

Traditional techniques such as statistical validation are effective and used for modeling the entirety or aspects of cognitive models. However, the more complex the cognitive model, the more ineffective these techniques. In statistical validation, a high correlation between actual and model data, or a strong *goodness of fit*, implies that the model is valid. However, one cannot conclusively state based on this heuristic alone that the model is fully validated, as high correlation may also be a result of overfitting, or where models fit both systematic and error variance within the same set [38]. Additionally, achieving such an amount of statistical accuracy involves focused, narrow, and limited models. While such models are critical in most scenarios, cognitive architectures are designed to allow for generalization, to be capable of multiple scenario applications. Kase et al. also points out that manual optimization, or fitting a model to human data, can be quite effective if modeling an individual or average performance across subjects, but requires significant computational resources [39]. In lieu of traditional manual optimization, Kase et al. proposed a new validation approach using genetic algorithms (GAs). With this approach, the GA is trying to determine the cognitive architecture parameters (i.e., independent variables) that best fits human data. In other words, the GA attempts to minimize the absolute difference in model and actual performance by varying cognitive architecture input parameters [39]. Since validation techniques require extra data in addition to the original dataset that informed the model, validation

is limited in small sample sizes. The validation scheme performed for the cognitive process model in this thesis is primarily focused on matching the simulation's chosen landing site with the landing site selected by a participant, assuming the same cue used as reported by the participant.

2.3 Crewed Spaceflight

A substantial amount of guidance exists for crewed spacecraft, with a primary focus on the health and sustainability of humans in extreme environments. These standards, or design guidelines, are captured in government documents (e.g. NASA-STD-3001 [40] [41], the Human Integration Design Handbook (HIDH) [42], the NASA Bioastronautics Databook [43], MIL-HDBK-761A [44], and MIL-STD-1472D [45]) and cover topics such as the human body capabilities, appropriate environments, habitats, hardware/equipment, and interfaces. Of these topics, the impact of human performance on trajectory design is of most relevance to this research. Already, aspects of this relationship are well-characterized: the constraints in induced acceleration, the range of viewing angles for ground observation by the crew, and the geometry of the Sun, Earth, and Moon for operable terrain contrast. Although much is known regarding the impact of environment and human performance on crew interface design, as evidenced by the presented government standards and literature regarding ecological interface design and user-interface design, these relationships are beyond the scope of this investigation.

The crew's role during descent and landing establishes direct constraints on several factors: the acceleration profile; the vehicle orientation and approach angle; and the launch window. The acceleration profile is primarily affected by the need to maintain human health, while providing sufficient reaction time and visual processing capability. The vehicle orientation and approach angle is driven by the dominance of the crew's role as a sensor and/or data interpretor. The launch window is also influenced by the responsibilities of the crew during data acquisition and processing. In addition to these direct constraints, general indirect constraints exist in order to achieve nominal human performance. Humans require minimum processing times for both audio and visual modalities and especially for

decision making. While these audio and visual modalities are known for most general cases, additional modeling is necessary to determine application-specific initial estimations on decision making time requirements.

2.3.1 Acceleration

Two sources of acceleration exist in spaceflight: those that occur naturally (i.e., Earth, Moon, Mars, etc), and those that are induced by the vehicle's movement (i.e., launch, landing, etc.). The human body is subject to both. In natural environments, the minimum acceleration that occurs is $0g$, or while in space, to $1g$, while on Earth's surface. Moon and Mars are fractions of Earth's gravity, at $0.17g$ and $0.38g$ respectively. The vehicle's trajectory, however, can impose even greater accelerations on the human body. For example, during launch, the crew may experience a maximum acceleration between 3-6g. During Earth re-entry, the peak decceleration may be even higher at 1.5-11.1g [42]. Spatial disorientation and other vestibular effects are the result of these changes in acceleration. Clark et al. analyzed this problem of astronaut spatial disorientation for several reference trajectories using a proposed Altair vehicle design [46]. The analysis illustrated significant differences in perceived and actual orientations, for example. Because of somatogravic illusions, the astronaut perceived a pitch angle of 0° when in reality, the vehicle was pitched forward 90°. The somatogravic illusion affects roll as well - the pilot commanded a roll angle of 45° but perceived an angle of 15°. The gross mis-estimation of vehicle angles has potentially negative consequences on pilot behavior. He may over-correct based on false perceptions and enter Pilot-Induced-Oscillation during landing, posing stability risks to the vehicle [46]. To minimize these risks, Clark et al. suggest careful consideration of pilot head placement within the vehicle.

The accelerations sustainable by astronauts are dependent on the type (linear, rotational), the duration (sustained: $\geq$ 0.5s, transient: $<$ 0.5s), and the direction with respect to the body (through the head, laterally, etc). Three axes are used to describe linear acceleration: G_x, G_y, and G_z. G_x is an axis parallel to the surface; directly through the chest, with the positive direction from the back to the chest ("eyeballs in", opposite is "eyeballs out").

G_y is also an axis parallel to the surface, but laterally positioned; the axis connects the two shoulders ("eyeballs left/right"). G_z is the axis perpendicular to the surface; directly through the head to the toe, with the positive direction from feet to head ("eyeballs down", opposite is "eyeballs up"). The body can withstand different accelerations at varying durations depending on the direction. Figure 1 illustrates the maximum limits for nominal conditions, assuming the astronaut has been conditioned to partial or excessive sustained accelerations.

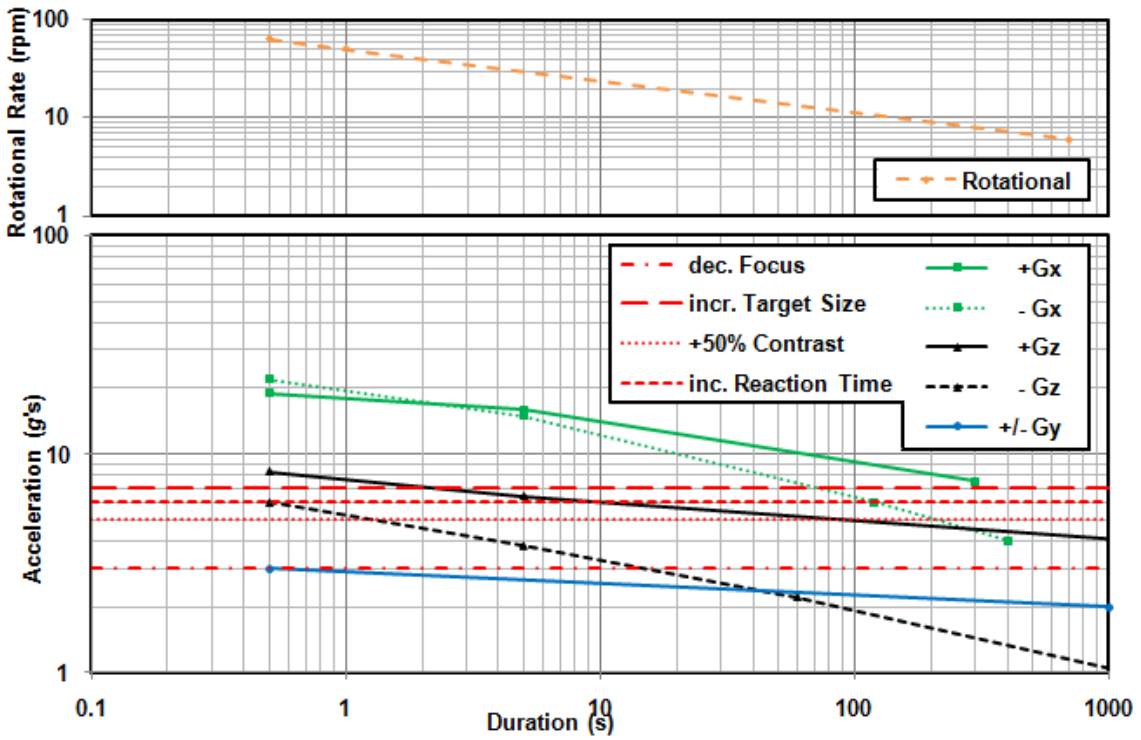


Figure 1: **Linear and Rotational Acceleration Limits.** Adapted from NASA-STD-3001 [41].

Further limitations may be placed on the vehicle trajectory and vehicle design if the pilot assumes additional responsibilities. Accelerations of large magnitudes or long durations affect several aspects of human performance: visual and audio perception capabilities and reaction time. At $+3G_x$, astronauts may have trouble focusing; at $+7G_x$, targets need to be twice as large in order to be seen [42]. Chambers and Hitchcock found two other benchmarks of human performance: at $+5G_x$, a 50% increase in contrast is required and at $+6G_x$, delayed reaction times occur. Thus, changes to the vehicle's trajectory may be

necessary, or the internal cockpit design must account for these human limitations once thresholds are crossed (e.g., displays, seats, etc).

2.3.2 Landing Site Visibility

Another defining attribute on automation design is the crew's role in perceiving the terrain. If the crew requires terrain visibility (assuming the lander has windows), then the vehicle must hold a specific orientation and approach relative to the targeted landing site. Otherwise, no additional constraints are imposed. The trajectory constraints are based on the position of the velocity with respect to the landing site and the vehicle configuration. Sostaric 2 defines two angles: the look angle and the depression angle. The look angle is defined as the angle between the vehicle's vertical axis and the line of sight to the landing point (a line connecting the center of mass and the site). The depression angle is measured between the horizontal and the line of sight, and is also known as the glide angle [1], descent trajectory elevation angle [47], or the visibility phase elevation angle [47]. Given the alignment of the velocity vector and the line of sight, this angle is also the same as the flight path angle. In his analysis of an Apollo-like reference vehicle, Sostaric determined that the limiting look angle was 25° (Figure 2).

The limiting look angle represents the case where the targeted landing site is at the edge of the window. At smaller angles, the landing site would not be visible to the crew. For reference, Apollo flew at a look angle of 51.9°. Intuitively, the optimal look angles for crew visibility and sensor operation are in direct conflict - more vertical approaches (decrease look angle) reduce the slant angle that increases sensor error and more horizontal approaches (incrase look angle) improve visibility for the crew. Additionally, the direction of landing site redesignation is constrained based on the approach. Vertical approaches allow for downrange diverts, whereas horizontal approaches improve reverse downrange, or uprange, diverts. The direction of landing site redesignation is dependent on the window configuration within the reference vehicle. Visibility of the landing site and alternative options may be improved with a different window geometry or location.

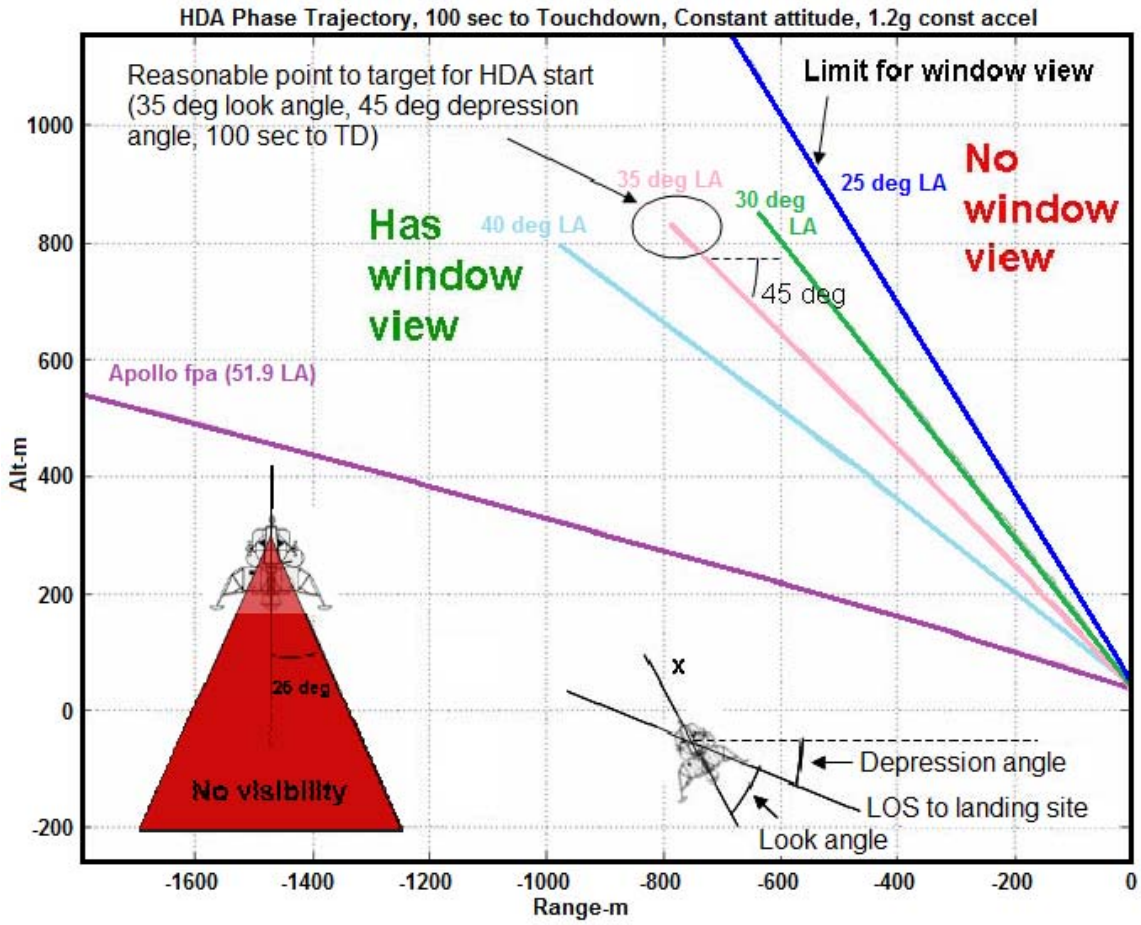


Figure 2: **Reference Approaches to Landing Site.** Adapted from Sostaric [48]

2.3.3 Lighting Effects

While landing site visibility can be improved by maneuvering the vehicle, visibility is also highly dependent on the terrain contrast, historically from the external lighting environment. Generally, the human eye can perceive objects within a range of 25-30 dB. Below this range, the eye sees darkness; above this range, the eye sees too much light. Within the range, the eye can discern differences of 0.1 dB. Stronger contrasts are preferred, particular during the landing sequence [1]. In order to meet this requirement, the sunlight must coming from a particular angle and direction. In preparing for the Apollo mission, NASA determined that the sun must be behind the vehicle (but not directly behind) [47] as to produce the necessary visible shadows for the crew. Additionally, the elevation angle had to be such that enough contrast could be discerned through the window without producing

Table 6: **Sun Angles and Flight Path Angles for each Apollo landing.**

Mission	Sun Angle Range [47]	Sun Angle Used	Flight Path Angle
11	5-13°	10.8°	16°
12		5.1°	
14			
15	7-23°	15°	25°
16			
17			

too much glare. Ideally, these objectives are met when the look angle is greater than the sun angle by more than two degrees in elevation or azimuth [47]. Contrast is worsened as the sun angle approaches the look angle [49]. In general for Apollo, the acceptable sun angle range is from 7-23°. Below this range, illumination may be difficult to achieve, as evidenced during Apollo 12 (terrain orientation difficulties). Above this range, the illumination is too great, and insufficient contrast is achieved [50]. Table 6 lists the sun angles used for the specific Apollo landings [1]. Achieving these sun angles meant waiting for the proper geometry between the Moon, Earth, and Sun, which occurred in one launch opportunity per month. The 7° range was selected primarily to allow for a 24-hour launch window.

An important consideration for landing on the South Pole is the lack of natural light. Recent studies such as the one conducted by Paschall et al. [51], suggest the use of external lighting, especially if the crew plays a significant role in landing site identification and terrain perception. The specific tradeoff between mass and external illumination is not explored in this thesis. Instead, the focus is on the effects of terrain contrast conditions on LPD performance, with three subsets - idealized, nominal, and poor - investigated in further detail.

2.3.4 Effects on Trajectory Design

NASA's Autonomous Landing and Hazard Avoidance Technology (ALHAT) team has also performed a trajectory trade during the ALHAT approach phase to understand the impact on the ability to detect hazards, the interactions of the crew with the automation, and the time to react to a hazard. Three parameters were varied: Flight Path Angle (discretized at 15 deg from 15-90 deg); six candidate Slant Ranges (0.5, 0.667, 0.8, 1, 1.5, 2km); and

the Engine Acceleration Profile (1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 2.0 lunar gs). Table 7 [51] summarizes the results of their findings.

Table 7: **Summary of Impact of Trajectory Parameters on Sensor, Crew, and Vehicle Capabilities.** This table is the results of a trade space parameter study performed by the ALHAT team and is adapted from Paschall et al. [51]

	Flight Path Angle	**Slant Range**	**Acceleration**
Sensor	Shallower angles are harder for sensors to scan site because of shadowing and pixel distortion	Closer slant ranges give better resolution but harder to obtain field of view	Lower accelerations provide more time for scanning and there is also less of an altitude change during a scan
Crew	Shallower angle allows for better view of the landing site out the window	Closer slant ranges give better resolution	Lower accelerations provide more time for scanning
Vehicle	Lower trajectory path angle have better nominal dV, but worse divert dV	Decisions made at higher slant ranges give more time to perform diverts which saves dV	Higher accelerations provide better dV characteristics

The trends noted by Paschall et al. also agree with the analysis of Cappellari in 1972, for the Apollo program. Cappellari calculated contours of constant ΔV cost for redesignations at 1200m (Figure 3). At this high up within the trajectory, reverse downrange diverts actually save propellant, as the time required to reach the site is shortened. The lower the redesignation occurs, the more propellant is required for any sort of divert, especially reverse downrange. In general, the mission was penalized by two kilograms of payload capacity for every additional kilogram of propellant alloted for redesignation [47].

Lastly, indirect constraints are imposed on the overall mission sequence due to minimal performance. Extensive testing has established minimal times for audio, visual, and information processing. The HIDH cites a number of studies for generally accepted response times. A college-age individual needs approximately 190 ms to react to visual stimuli, particularly light [42]. Auditory stimuli require slightly more processing time, at 160 ms. The stimuli will be registered in short-term memory for an approximate period of time. To avoid superimposing two different chunks, the visual register requires 300 ms for clearance [42], whereas auditory requires 20000 ms (20 s) [42]. Other reaction response times have been recorded beyond the auditory and visual modalities. Welford determined that 93 ms are

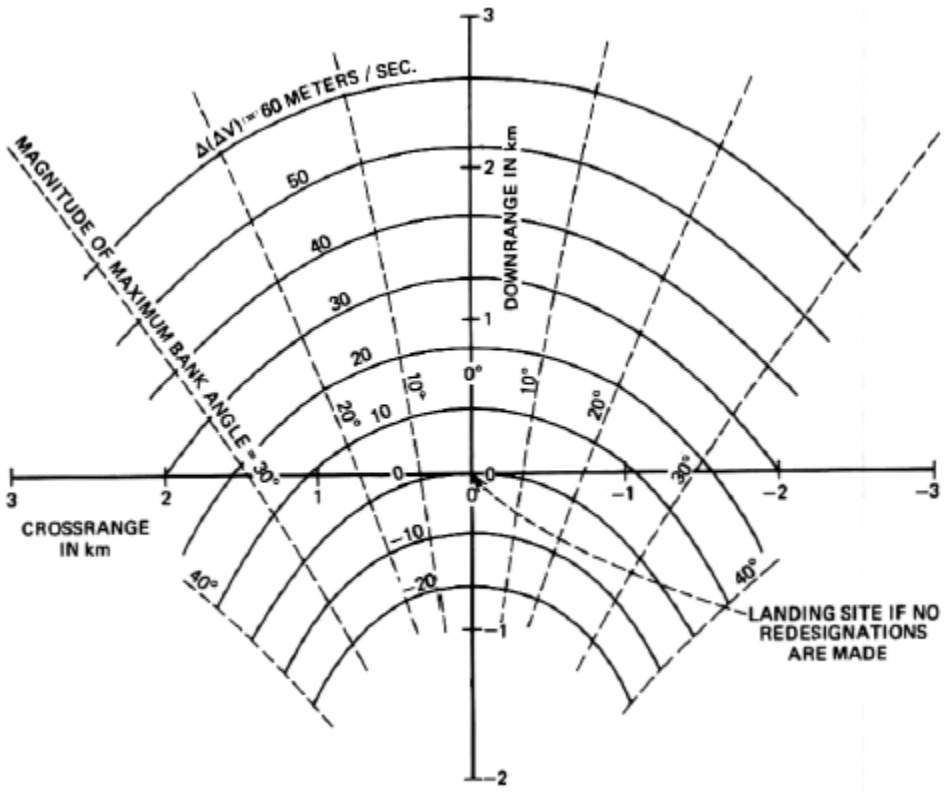


Figure 3: **Constant ΔV penalty for landing site redesignations at 1200m altitude.** Reproduced from Cappellari [47]

.

required for each inspection [52]. Cavanaugh found a series of rates at which items could be matched against memorized objects: digits (27-39 ms/item), colors (38 ms/item), letters (24-65 ms/item), and words (36-52 ms/item) [53]. Additionally, Card et al. developed a methodology for estimating task completion times for standard computer interactions. While astronauts are not expected to interact with a keyboard, mouse, and computer monitor, there may be display interfaces involving button presses (500 ms); moving a joystick (1100 ms); or mentally prepare (1350 ms) [34].

2.4 Related Studies

Current literature presents a number of studies related to function allocation for LPD, evaluation of specific functions or displays for LPD, and the first instances of LPD during the Apollo program. Additionally, NASA has revisited the problem of landing on the Moon and attempted to improve the landing technology. The ALHAT project is focused

on leveraging technological advancements in terrain sensing and alleviating crew workload. Much of the work described in Chapter 3 is derived from this literature.

2.4.1 Apollo

After almost ten years of development work, three days in transit to the Moon, the fledgling United States space program finally made the first visit to the Moon in 1969. However, the voyage from lunar orbit to the ground was far from epic: a diversion to avoid terrain obstacles and the threat of fuel exhaustion. The Lunar Excursion Module (LEM) separated from the Command Service Module in lunar orbit and embarked on a 30 minute traverse [54], with the intention of landing at predesignated landing areas. The Apollo astronauts performed LPD at what is considered, by today's standards, a low level of automation. However, for the men who were test-pilots-converted-to-astronauts, they felt like "spam in a can" [55], delegated to being a passive component of the vehicle. The astronauts' feelings were best summarized by John Glenn: "We don't want to just sit there and be just like a passenger aboard this thing. We will be working the controls [56]." Glenn's mentality was most likely a result of the test pilot culture and the desire to retain control in a potentially life-threatening situation. The risks and uncertainties associated with this new mission demanded the development of a digital autopilot. Thus, biological and digital pilot shared responsibilities during descent, especially during LPD.

LPD began after Powered Descent Initiation (PDI) and shortly after entering "high gate" (Program 64/P64) [57][58]. At high gate, the LEM was at a range of 26000 ft [59], an altitude of 7515 ft, had a descent rate of -145 fps, and an inertial velocity of 506 fps [58]. The terrain was visible to the crew during this approach phase. The vehicle entered "low gate" (P65/P66) at a range and altitude of 2000 ft [59] and 500 ft, respectively, with a descent rate of 16 fps, and an inertial velocity of 55 fps. The terms high and low gate were derived from aircraft pilot terminology, particularly for approach to an airport [58]. During this time, the LEM flew at a flight path angle of 16° for Apollo 11-14, and 25° for the later missions [47]. This orientation allowed astronauts to view the lunar terrain from the LEM window, a requirement to land on the Moon, based on the allocated responsibilities.

After performing LPD, or several instances of redesignation, the pilots, with automated assistance, performed a vertical touchdown. Overall, Apollo had a landing ellipse of about 20km downrange and 5km crossrange [51], with terminal conditions listed in Table 8.

Table 8: **Apollo LEM Touchdown Limitations.** Replicated from Paschall et al. [51]

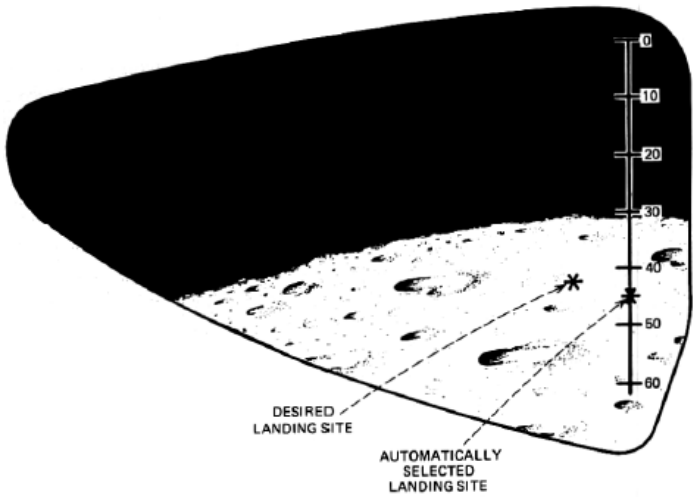
Vehicle State	**Value**
Vertical Velocity V_V	10 ft/s
Horizontal Velocity V_h	For $V_V \leq 7$ ft/s, $V_h = 4$ ft/s For $7 \leq V_V \leq 10$ ft/s, $V_h = \frac{40}{3} - \frac{4}{3}V_V$ ft/s

.

The Apollo astronauts performed LPD at a relatively high degree of control, relying predominantly on their perception of the lunar terrain as seen through the LEM window. Both the Commander (who piloted the LEM) and the Lunar Module Pilot (LMP) (who operated the other space systems and informed the Commander of vehicle and mission status) [60] participated in LPD, with very specific roles. The Commander worked primarily with the landing point designator (Figure 4). The landing point designator is a reticle-etched window located on the left side of the LEM. The Commander would align the reticles etched on the outer and inner windows and view the lunar terrain across 2° (vertical) and 5° (horizontal) scales. A representative example of this terrain is presented in Figure 5. The digital autopilot would indicate where to find the designated landing site, which the LMP would then read to the Commander. If the Commander opted to land at an alternative location, he would call in this decision by moving the control stick in the direction and number of "clicks" as proportionate to the landing point designator. Nudges fore or aft would move the landing site up or down (relative to the Commander's field of view) by 0.5°. Lateral nudges would shift the landing site by 2° [55] [58]. This process continued for several iterations, with the LMP reading until the LEM reached the predesignated time-to-go and performed terminal descent.

In general, the pilots were tasked with collecting and interpreting information regarding the lunar terrain, creating landing site options, selecting the option that, given current information, was most appropriate, and executing the maneuvers necessary to bring the vehicle to the final destination. In parallel, the digital autopilot was also collecting (or

(a) Landing Point Designator [54].



(b) Schematic of Operation [47].

Figure 4: **Landing Point Designation for Apollo.**

sensing) the lunar terrain and executing aspects of the vehicle maneuvers. Major et al. analyzed the Apollo transcripts and determined that the crew spent 50% of the time during PDI (and before touchdown) communicating with mission control. This time was spent discussing: vehicle position (43%); vehicle status (27%); mission schedule (17%); and other topics (13%) [61]. In general, the crew was concerned with monitoring, diagnosing, scheduling, and terrain assessment. Terrain assessment proved to be a particular challenge to the crew, with a number of redesignations occurring for most of the Apollo landings. Table 9 summaries these redesignations and the range displacements from the initially targeted site.

Table 9: **Number of Redesignations and Range Displacements during Apollo.**

Mission	**Number of Redesignations (in P64)**	**Range Redesignations displaced LM from Landing Site**
Apollo 11	Switch to P66 early to avoid boulder field	–
Apollo 12	7	–
Apollo 14	1	2000 ft downrange, 300 ft north
Apollo 15	18	1110 ft uprange, 1341 ft north
Apollo 16	10	620 ft uprange, 635 ft south
Apollo 17	8	–

The digital autopilot was tasked with processing data from the radar and orienting the LEM to ensure clear communication channels with Mission Control. During LPD,

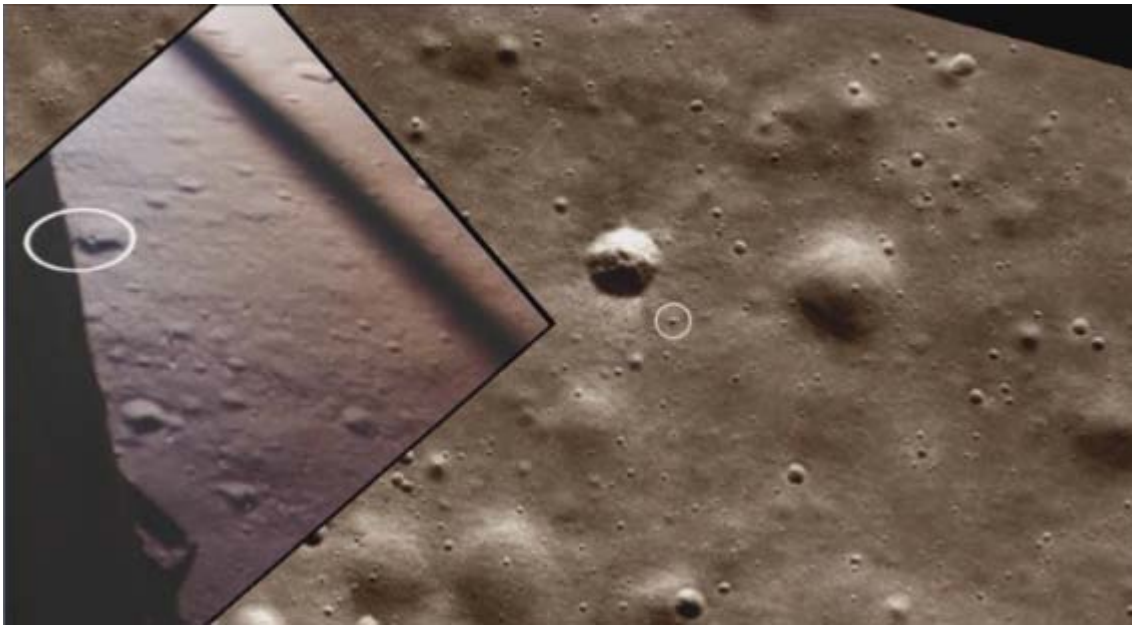


Figure 5: **Out-the-Window View for Apollo 11.** The image is a screenshot taken from the film "The Final Approach" by Cantin and Cantin [62]. The image on the left is from the Commander's window; the image on the right is an top-down map of the lunar surface. At this instant, the LEM is 350 ft from the surface, a downward velocity of 9 ft/s, moving forward at 58 ft/s.

the autopilot was also loaded with a series of guidance computer programs, including the *Landing Maneuver Approach Phase* or Program 64 (P64) and the *Velocity Nulling Guidance* or Program 65 (P65). The crew also had optional programs, such as P66, *Rate of Descent*, or P67, *Manual Guidance*. P66 had several options: the flight computer controls the vertical speed and nulls the horizontal speed; the flight computer controls only the vertical speed and the crew controls the final attitude; the flight computer nulls the horizontal speed while the crew controls the engine throttle (*i.e.*, the descent rate); the crew controls the rate of descent and the attitude and lands the LEM themselves [59]. P67 was only to be used if P66 failed and permitted complete crew control of the engine throttle. All Apollo landings were flown using P66, with the crew dialing in this program before the automatic switch to P65. To interface with the digital autopilot, the crew used the Display and Keyboard unit (DSKY), to dial in two-digit numbers that represented verbs and nouns.

Tasking the crew with such a large responsibility during landing has interesting consequences. The crew can adapt to unforeseen situations, such as the 1202 alarm that kept

misfiring during the Apollo 11 landing. Additionally, the crew can correct for accumulating errors, such as Commander Neil Armstrong taking over manual control to avoid a rough area of terrain. However, the crew may be unpredictable, and deviate from initial plans based on their own preferences. Figure 6 illustrates altitude/altitude-rate profile for Apollo 11. As illustrated in this figure, there are significant deviances between the actual and predicted profiles. When asked, Armstrong responded that he was "just absolutely adamant about my God-given right to be wishy-washy about where I was going to land" [58]. While these types of maneuvers are necessary, the pilot must calculate the tradeoffs between achieving the mission, landing in a safe region, and fuel expenditures (another contribution to mission safety).

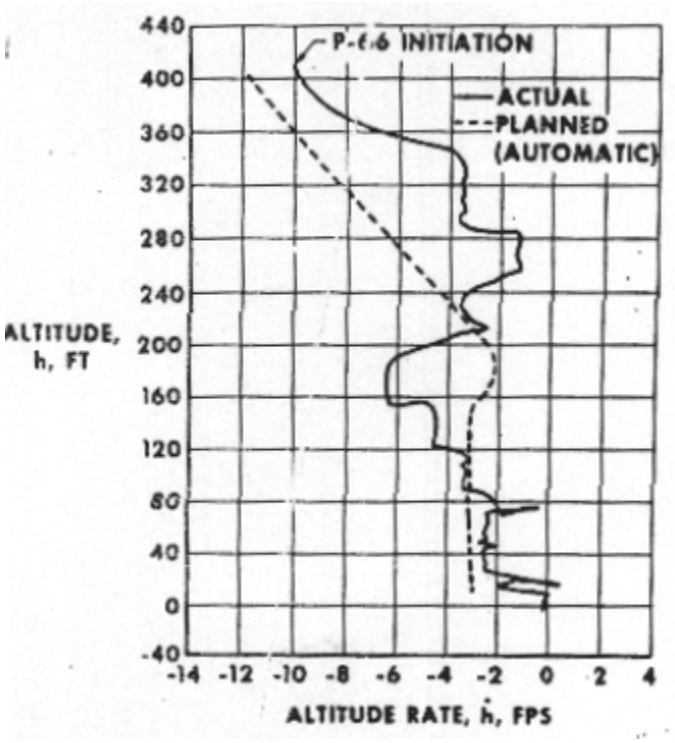


Figure 6: **Altitude/Altitude rate profile for Apollo 11**. Replicated from Bennett [58].

2.4.2 Autonomous Landing and Hazard Avoidance Technology

The ALHAT project was formed in 2006 in response to the growing need for improved landing technology. The mission statement for the ALHAT project is to "develop and mature to technology readiness level 6 an autonomous lunar landing GNC and sensing

system for crewed, cargo, and robotic lunar descent vehicles. The ALHAT System will be capable of identifying and avoiding surface hazards to enable a safe precision landing to within tens of meters of certified and designated landing sites anywhere on the Moon under any lighting conditions" [63]. The ALHAT system itself consists of a sensor package, an autonomous flight manager (AFM), and a human-systems interface for supervisory control. ALHAT is operational during low lunar orbit, the transfer orbit to the surface, and powered descent - the later of which contains LPD. LPD is referred to as Landing Point *Re*designation (LPR) by the ALHAT team as a change in landing site destination is redesignation from the pre-selected baseline site. This thesis uses Designation, as a baseline site is not necessarily in place.

The powered descent mission segment takes approximately 7-14 minutes to complete, with LPR lasting about 30-120s. Initiation of powered descent is selected so the respective trajectory meets a targeted altitude and range from the landing site, given the specific thrust-to-weight ratio of the landing vehicle. Once the braking burn is initiated, the vehicle will perform a pitch-up maneuver (approximately 1km altitude) to allow for sensor and human visibility of the landing site, approach the site at a fixed orientation to allow for landing point redesignation, and then perform terminal descent (about 30m altitude) [64] once a final site is selected. Nominally, the approach phase begins at a slant range and angle of 1029m and 44.9° respectively. These initial conditions leave about 67s of time available for hazard detection and avoidance, with about 695m of available altitude for diverts [65].

During this approach phase, ALHAT conducts LPR similar to Apollo: terrain sensing, data processing, alternative site planning, site selection decision making, and vehicle execution. However, the main difference between ALHAT LPR and Apollo LPR is the distribution of authority between the crew and the AFM. ALHAT operates under a much higher level of automation, relegating the crew to a supervisory role. The AFM performs hazard detection and avoidance by using a LIght Detection and RADAR (LIDAR) sensor to map the terrain (approximately 5s). The LIDAR data is then processed and the AFM produces a map of the landing area with hazardous areas indicated (approximately another 5s). Additionally, a cost map is generated based on vehicle safety, fuel consumption, and/or

points of interest and three regions of favorable costs are selected as alternative sites. These sites are presented to the crew, who are then able to choose an alternative site within the alloted time (30s). Once the crew makes a final decision, the AFM reassigns the new target conditions to the guidance system, and automatically flies the vehicle to the target site, concluding with terminal vertical descent.

Overall, the crew performs more of a managerial role within the ALHAT framework, in lieu of being an integral and crucial component of the system. The crew monitors the ALHAT system and checks for abnormalities. If a change is necessary, the crew may "reach-into" the AFM system and make necessary conditions. In particular, target conditions may be adjusted to satisfy the crew's needs [66]. For example, during LPR, the crew's role is no longer conducting various subsystems and piloting the vehicle, but, rather, is determining the best landing site of three alternatives.

Regardless of whether crew is present or not, the AFM is part of the ALHAT System and thus, always present. Many of the traditional crew roles have now been assigned to the AFM, such as supervising the GNC software and sensor situations. The ALHAT system is uploaded with sets of target conditions prior to launch. During the mission, the AFM monitors and ensures these target conditions are met while permitting a safe landing plan [66]. To fulfill these roles, the AFM communicates with several sensors. The ALHAT sensor suite consists of an inertial measurement unit (IMU), a star tracker, a laser altimeter, a flash LIDAR sensor (which acts as the hazard detection and avoidance sensor), and a Doppler LIDAR sensor [64]. The sensors are operable at different altitude ranges, the IMU having the largest range to the flash LIDAR with the shortest range. The IMU is operable from 100km altitude and through the rest of the trajectory; the altimeter can be initiated at about 15km to ground; the Doppler LIDAR velocimeter at 2km to ground; and the HDA sensor begins operation at approximately 1km slant range from the landing site.

2.4.3 Function Allocation and Evaluation for LPD

Several key studies have focused on various elements of LPR. Forest, et al. developed a landing site selection algorithm that, when given terrain data, would highlight key hazards

and suggest alternative sites based on the cost function preference of the crew. This study provided an initial reference LPR display, but did not model human interaction with such a system [67]. Needham investigated the impact of varying levels of automation on human performance during LPR, concluding that higher automation allowed for quicker time to complete [68]. In addition, Needham developed a set of icons that would overlay landing site terrain characteristics on a top-down synthetic map. An experiment was also performed to observe the impact of varying levels of automation. However, the subjects used in this experiment were graduate students with little piloting experience and not closely representative of astronauts. Wen et al. [69, 70] also performed a similar study on LPD and included the task of piloting to touchdown. These results include comments on system performance and variability of measures such as fuel usage and landing accuracy. This study was completed with students and examines one function allocation. Lastly, Chua, et al. have derived a task model and used this model to examine bottlenecks of LPR [71]. These bottlenecks were addressed by redesigning the LPR display to simplify the information layout and to utilize new symbolism to represent site characteristics. This LPR task model also incorporates expert decision making theory [31] to account for specialized astronaut behavior [72]. However, this study is based on theory and lacks observations from equatable subjects. Duda et al. [73] have also designed lunar lander manual control modes that are analogous to, but improvements on, the original Apollo control modes. Additionally, several evaluations have been conducted with the ALHAT team regarding interface usage and display design [74] [75, 76]. The results have evaluated the use of synthetic vision or different viewpoints and suggested symbology for efficient information gathering and processing. Major et al. [77] have also outlined an approach for the design and testing of technology developed to support crew tasks during space exploration.

CHAPTER III

DECISION MAKING DURING LANDING POINT DESIGNATION

Landing Point Designation (LPD) is a critical landing task that poses a unique challenge to mission designers. During this task, the vehicle performs a pitching maneuver, operates sensors to scan the expected landing area, and chooses a final landing site, factoring in criteria such as fuel consumption, site safety, and proximity to points of interest (*e.g.,* scientific phenomena or previously landed assets). Mission designers have two main issues to address when designing for LPD: the allocation of functions between crew and automation; and the information and environment requirements associated with different allocations. In order to logically allocate functions and therefore design the most appropriate automation system for these tasks, mission designers need to understand the crew's judgment and decision making process, the capabilities of the automation, and the information requirements for both crew and automation. Ascertaining this information requires decomposition of the LPD task and to determine the necessary information to develop a cognitive process model of human-automation decision making. Since evaluating the full spectrum of function allocation combinations is impractical, this thesis focuses on a smaller subset of cases - three representations of potential function allocations. The crew is modeled as a single astronaut, but may consist of two or more astronauts working together. This chapter discusses the task decomposition, function allocations examined in this thesis, and a series of human-in-the-loop (HITL) experiments conducted to characterize the human-automation interaction during LPD.

3.1 Task Decomposition for Landing Point Designation

All tasks are composed of several fundamental subtasks, such as sensing, interpreting, creating, selecting, executing. Specifically within LPD, the agent (human, automation) senses information through the vehicle window in addition to, or in lieu of, sensing the terrain with a LIght Detection And Radar (LIDAR) sensor; the agent interprets the sensed data

and assesses the situation and data significance; the agent creates landing site options; the agent selects an appropriate landing site; and then the agent commands the vehicle to the chosen landing site. Generally, there are three different options for completing each of the subtasks: human does all; automation does all; or human and automation share the subtask. Therefore, 243 combinations of completing LPD exist. Three of these combinations were chosen as candidate function allocations and described in Table 10: the Apollo-like, moderate, and robotic cases. Performance is expected to change dependent on the owner of the task. To provide context to a known function allocation framework, the corresponding Endsley and Kaber [13] ten-point levels of automation are approximately Level 3 for the Apollo-like function allocation, Level 7 for moderate; and Level 10 for robotic.

Table 10: **LPD Task Decomposition and Function Allocation.**

Allocation	**Sensing**	**Interpreting**	**Creating**	**Selecting**	**Executing**
All Crew	Human	Human	Human	Human	Human
Apollo-like	Human/Auto	Human/Auto	Human	Human	Auto
Moderate	Human/Auto	Human/Auto	Auto	Human	Auto
Robotic	Auto	Auto	Auto	Auto	Auto

Using this framework and current literature, one can use cognitive task analysis to develop a model of LPD. Figure 7 illustrates the three function allocations and the tasks completed by human or automation. Each processing cycle is iterative. As the vehicle approaches the ground, the human or automation may continue to receive information regarding the landing area, therefore repeating the following subtasks as necessary. Aspects of the sequence may be iterative, such as within the Apollo-like function allocation with creating and selecting sites and continuing to refine that final decision. Astronauts may also skip steps within this sequence. If the results of the sensor scan match expectations, then the astronaut immediately selects a site without further consideration. These expectations are set prior to the task, a cognitive phenomenon known as a mental model. Mental models are the models held by the human regarding the task. In this thesis, the mental model is assumed to be an expectation of where to land within a given landing location. Training is assumed to establish this particular mental model, although environmental factors and user experience are also influences.

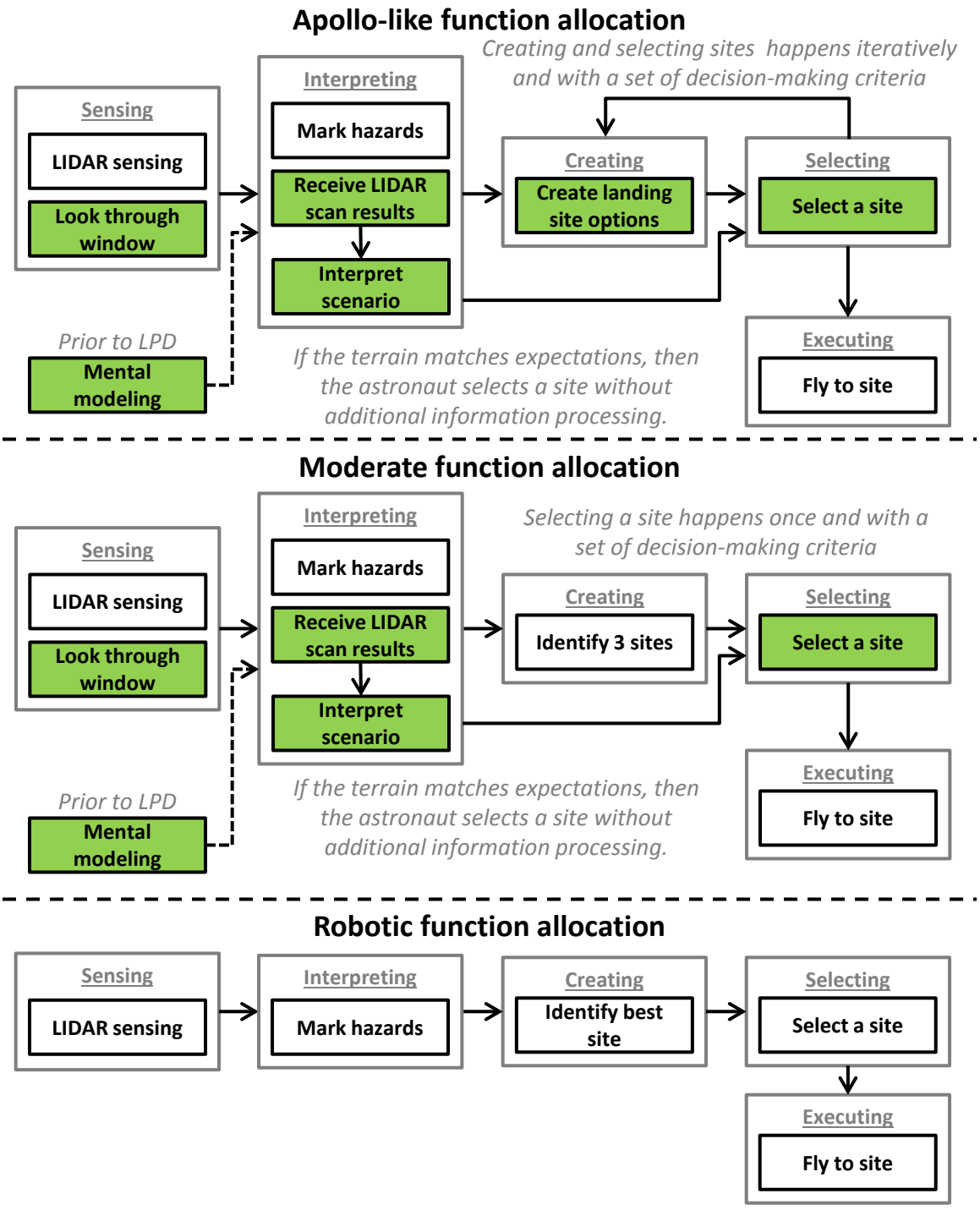


Figure 7: **Task analysis for Apollo-like, moderate, and robotic function allocations.** Gray boxes are fundamental tasks that can be completed by a human or a computer. White boxes are tasks completed by the reference automation. Green boxes are tasks completed by the human.

There are several key differences between each of the function allocations. Overall, the LPD task is expected to be completed fastest in the robotic case, finishing the task in 5s. The site chosen is always the optimal site. With the crew relegated to payload status, the vehicle is less constrained by viewing angles. The only major considerations on the trajectory are optimizing fuel consumption and ensuring the best approach for sensor operation. For purposes of equal comparison, the robotic case is assumed to fly the same trajectory prior to LPD as the other two cases and begins the task at the same initial conditions outlined for the Apollo-like and moderate cases. The LPD algorithm in this case places equal weighting on fuel and safety, unlike the other cases, where the crew may prefer one criterion over another. Second, the humans in the moderate function allocation are not permitted to suggest a landing site or to revise their final decision. Once a landing site is chosen, it is finalized. This function allocation simplifies the LPD task for the user by limiting the search region to the top landing sites. The reference automation in both the robotic and moderate function allocations choose sites based on priorities established by mission designers and coded into the automation. Lastly, the Apollo-like function allocation permits redesignation, or multiple opportunities to select or adjust the chosen landing site. The automation in all three function allocations pilots the vehicle to each designated site until terminal touchdown conditions are met. Given the limited time, high stakes, and highly rehearsed nature of the task, under nominal conditions, it is more likely that the crew will settle on one landing site that has proven itself during training, than generate a series of alternatives and cross-compare. However, as of this thesis, the appropriate training for LPD is unknown.

What is still unknown about this task is the specific interaction between the crew and the Apollo-like and moderate automation systems, in particular, the decision making strategy used for both. This strategy is composed of the criteria used to create and select a final landing site, the priorities associated with those criteria, and the mechanisms for finding the landing site. Additionally, it also unknown as to how these strategies change with respect to different function allocations. As no humans are involved with the robotic case, these changes are only captured in the Apollo-like and moderate function allocations. Current

literature does not provide enough data to support these questions and observing the LPD task is not practical. Therefore, HITL experiments were needed to characterize the human-automation interaction that occurs during LPD, provide data for the development of a cognitive model, and ultimately derive system requirements for automation system design. These specific elements - which decision making criteria, the number of criteria, the relative importance, the site search method - are useful in determining which aspects of the task or the environment need to be presented on displays, or brought to the crew's attention. Effectively, these criteria are piece of information needed by the crew to complete LPD.

Two experiments were conducted specifically for this thesis, based on the results of an earlier experiment. All three experiments were designed to further understanding of human decision making during LPD, but with slight variations associated with the reference automation and independent variables. Table 11 illustrates the differences between each of these experiments. Summaries of the experiment designs and results of Experiments 1 and 2 are presented in 3.2 and 3.3. The design of experiment 3, or the main thesis experiment, is explained in detail in Section 3.4. The results are discussed in Chapter 4. For a more thorough discussion on Experiment 1, see Chua, Feigh, and Braun [78, 79] and Chua and Feigh [80]. For further information on Experiment 2, Chua and Feigh [81] reports the design and results.

3.2 Experiment 1

The first experiment in this series was conducted in 2009. This experiment involved nineteen fixed-wing pilots, who performed the LPD task in a simulated lunar lander cockpit at Georgia Tech. The reference automation was considered moderate and the functionality was the same as defined in Section 3.1, but this automation was not used in any of the other studies. The system that featured "hot keys", a series of buttons which allow the crew to choose a predesignated weight distribution for the landing objective function. Three sites would be suggested with each hot key. Pilots needed to land near one or two points of interest (POIs) and saw either 1, 3 or 2, 4 identifiable terrain markers. Additionally, some of the runs featured terrain that differed from what the pilots were prepared for, to account

Table 11: **Experiment sequence and differences.**

	Experiment 1	**Experiment 2**	**Main**
When	2009	2011	2012
Facility	Georgia Tech	Georgia Tech	NASA JSC, ROC [1]
Demographics	Fixed wing pilots	Helicopter pilots	Astronaut Office
Number	19	15	13
Function Allocation	Moderate [2]	Moderate	Apollo-like, Moderate
Independent variables			
Terrain markers	x		
Point of interest	x		
Terrain expectancy	x		
Lighting		x	
Scenario			x
Automation			x
Trajectory			x
Dependent variables			
Pilot performance score	x	x	x
Landing Site Score		x	x
Time to complete	x	x	x
Situation awareness	x		
Strategy: criteria	x	x	x
Strategy: preferences	x	x	x
Strategy: search method		x	x
Strategy: sequence	x	x	x
Workload			x

for changes in terrain expectancy.

Overall, the experiment results showed that none of the independent variables had any significant effect on the quantitatively measured dependent variables. The analysis and observations of the debriefing session after each participant, however, illuminated a number of key insights to the decision making process. First, pilots commented on generating a mental map of the expected landing area and mentally highlighting favored "sweet spots", "zones", or "quadrants" of where ideal landing sites were thought to manifest. This mental map was created pre-scenario briefing (represented as satellite photography) provided before every run, or during initial evaluation of the LIDAR sensor scan results. This mental modeling exercise was thought to occur during terrain orientation only, not specifically for site selection. The discussion of these highlighted areas also led to the hypothesis that participants narrowed choices by eliminating sites that did not belong in these areas.

Second, pilots illustrated two different strategies for selecting a landing site. Several pilots reported "sharing" responsibilities with the reference automation. These pilots belonged to the single/double-button party, preferring to optimize on only one landing metric (fuel, safety, proximity to POI), and expecting the automation to optimize on the other qualities. Pilots that employed the multiple objective function strategy also utilized various landing site search methods. Several pilots treated the hot key buttons as a binning system rather than a filtering mechanism. As one pilot stated, "[the researchers] could've labeled [the buttons] as anything, I didn't pay much attention to what they actually stood for". Generally, these pilots found the hot key arrangement to be impeditive - several pilots suggested a "see all button"; a method of flagging sites for further investigation; or a means to eliminate undesirable options. Although workload was not measured in this experiment, these pilots commented that recalling each site and the affiliated hot key, was a challenge.

Even the order of hot key use was varied. Some pilots used every hot key to quickly scan the site alternatives; hot keys containing sites of interest would be returned to for further investigation. Other pilots engaged the hot keys, as intended, based on their site criteria priority. For example, one pilot regarded the order of safety, fuel, and then proximity to POI as his preference of priority (greatest to least), and the order of hot key use reflected

this preference. Other pilots engaged the hot keys in the same order they were presented, but with a purpose. One pilot believed that more critical objective functions should be physically closer to the pilot. He also felt that a cockpit interface designer should arrange the hot keys in a similar fashion, so pilots who were new to the vehicle (he regarded himself as such a pilot) would comprehend that this construction was the order of objective function importance. The pilot's attitude, while most likely not universal, reinforces the significance of the cockpit display interface. Some pilots attempted to circumvent the hot keys. Several pilots engaged only the balanced or *a priori* hot keys and further optimized within those options on one criterion. Most selected a site that required the least amount of fuel and consistently made this decision quickly. One pilot explained he was more confident in his abilities to optimize fuel consumption than the automation. Another pilot stated a desire to maximize his control over this complicated task by restricting his options to a select few.

Experiment 1 established the foundation on which Experiment 2 was built. First, the use of terrain markers (e.g., hazards) was determined to be an ineffective measure of terrain difficulty. Observations showed that this representation of distinct items of interest was not universal between participants. The map may not have established an appreciable distinction between low (1,2) or high (3,4) numbers of hazards. Second, participants mentally modeled where sites would be, but how those mental models were developed was unclear. Additionally, this concept of discarding "bad" zones (and sites that appeared in those bad zones) hinted at different decision strategies. Lastly, during debriefing sessions with participants, it became clear that more than five criteria (corresponding to the hot keys) were used during the decision making process. Subsequently, another experiment was necessary to provide more details on these findings.

3.3 Experiment 2 Overview

The second experiment was conducted in 2012. This experiment involved fifteen helicopter pilots, who performed the LPD task in the same simulated lunar lander cockpit as the previous experiment, but with a different reference automation. The automation was considered to be a moderate function allocation, but the functionality differed from the first

experiment. The system did not use "hot keys", but instead provided three landing site options to the pilot, based on safety and fuel consumption criteria. This functionality is the same used in the main experiment. Pilots were asked to choose one of the three sites in two lighting conditions: poor, ideal. Quantitative measures such as landing site score, pilot performance score (PPS), and time to complete were recorded and a debriefing session was conducted with each pilot. The PPS formulation differed in Experiment 2 from Experiment 1.

Statistical analysis indicated that there was no significant effect due to lighting on the dependent variables, therefore establishing the lighting conditions that LPD performance is reasonably consistent. The qualitative analysis, however, reaffirmed the initial observations made in Experiment 1. With moderate automation, participants were observed to belong in one of two types of decision strategies: reranking or eliminating. This framework was used in the Main experiment and definitions are provided in Chapter 4. Follow-up analysis demonstrated that neither strategy provided a significant advantage in overall LPD performance. Questions on mental modeling were also posed to the participants. If used, participants were asked to annotate a map of the expected landing area, marking the areas of what they considered to be good and bad. Statistical analysis showed that participants who performed mental modeling significantly decreased the time to complete, but the performance was not significantly altered.

In addition to inquiring the overall decision making strategy, participants were polled on the criteria used and the relative weighting associated with each criterion. Nine criteria, or cues, prompted actions (preferences for sites) from the participants. Additionally, the relative importance of these cues was assessed, with participants asked to provide a numerical weighting or a ranking of the cues. Participants were also asked for definitions on vague descriptors such as large, buffer, good, bad. This type of data is useful for establishing a rule-based cognitive process model of decision making for the LPD task. A prototypical model was developed based on these experiment results, with integrating the cues, the weightings, and codifying the cognitive actions and decisions made by participants. Initial modeling results showed reasonable model agreement with experiment results.

Experiment 2 furthered understanding of decision making during LPD and established an effective pilot study prior to the main thesis experiment. Many of the qualitative analyses conducted in this experiment were refined and used again. An effective range of lighting conditions was determined, allowing for the main experiment to focus on one lighting type. Additionally, comments made during the debriefing helped determine which characteristics differentiated terrain. Figure 11 illustrates these characteristics: perceived hazards (based on Gestalt principles) and shadowing.

Understanding the types of search methods possible with a moderate level of automation allows for contextual framing of questions to further elucidate the decision making options. However, Apollo-like automation was not tested in this experiment, nor were there variations to the environment. Experiments 1 and 2 also focused on nominal operations. No data existed to examine the potential changes due to off-nominal scenarios, or a comparison of human-automation interaction. Furthermore, Experiment 1 and 2 focused on fixed-wing or helicopter pilots. It was not known if astronauts would exhibit different behavior due to their specialized training.

3.4 Main Experiment Overview

The main purpose of this experiment was to understand how astronauts make decisions during LPD, especially when interacting with an automated decision making aid. Pilots and commanders were the targeted demographic, however, due to the limited availability of these experts and the possibility of mission specialists engaging in such activities, all astronauts were evaluated in this study. It was hypothesized that the astronauts would use different strategies to complete the LPD in response to the *role of automation*. Other factors such as *landing trajectory* or *scenario type* are likely to prompt different strategy use, but it is unclear the amount of change. The extremes of each independent variable was tested: landing trajectory (shallow, baseline); scenario type (nominal, LIDAR sensor warning); and function allocation (moderate, Apollo-like). Landing trajectory and scenario type were used in this experiment as within-subject independent variables. Limitations in resources resulted in the function allocation used as a between-subjects variable. Therefore, each

participant saw four of the eight possible scenarios. The order of the runs within-subjects was balanced to reduce any potential bias in run order.

The baseline and shallow landing trajectories were chosen to encapsulate the range of possible LPD times presented to the onboard crew. The baseline trajectory assumes a lunar acceleration limit of 1.1 lunar-g's beginning at a slant range of 1000m and a flight path angle of 30°, providing a 50.4° look angle. These conditions provide 78s prior to the final vertical descent maneuver. The shallow trajectory, assuming the same acceleration limit, begins at 500m slant range and 15° flight path angle, resulting in a 45s for the LPD task. Figure 8 illustrates the differences between the two trajectories.

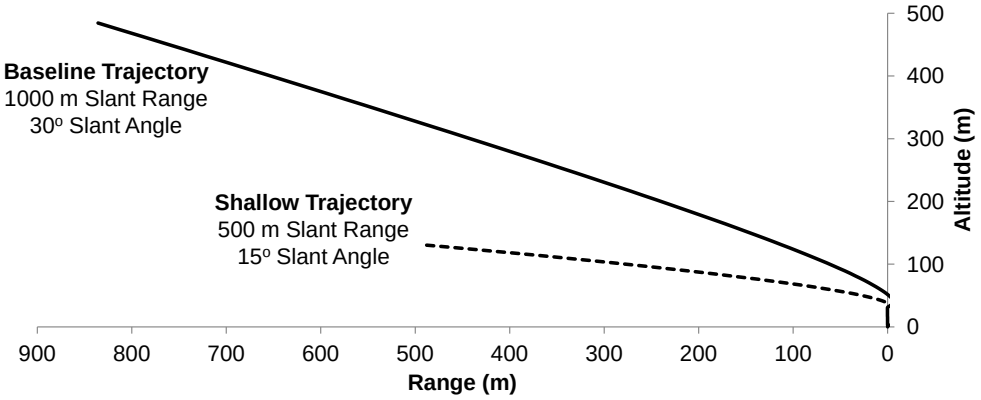


Figure 8: **Baseline and shallow trajectories.**

Under nominal conditions, the automation, sensors, and other systems perform without error. Most automation systems are designed and work well in these circumstances. However, as illustrated in several commercial aviation incidents and notable spacecraft flights (e.g., Apollo 11 and 12 landings [3]), the unexpected may occur and require human intervention. A malfunctioning sensor was chosen to represent a feasible scenario that exceeds the nominal bounds of automation performance. The LIDAR scan is taken of another site different from what the crew is expecting. The sensor performance scenario represents a

[3] 1201 guidance alarm almost caused an abort on Apollo 11 and slosh issues persistently disrupted guidance calculations on Apollo 12.

damaged LIDAR sensor. The cause for this degraded performance is not specified, the participants were told the LIDAR scan may misidentify hazardous and non-hazardous regions.

The function allocations were selected to represent two potential automation systems designed for different responsibilities of the human. The moderate function allocation proposed in this thesis is equivalent to the ALHAT design. Three landing sites are suggested to the crew by the reference automation system in addition to identification of hazardous sites. The participants can only select once and cannot suggest an alternative site. In contrast, the Apollo-like function allocation represents a technological improvement over the Apollo era. No sites are suggested, but the identified hazards are presented to the crew. The crew must choose where to land and are able to modify this decision through the end of LPD. The reference automation system for both function allocations pilots the vehicle to the end of LPD.

Five dependent measures were collected and two derived from collected measures. *Time to complete* (s) was measured by the start of the LPD task and the time elapsed until the participant selected a landing site. In the case of the low automation, where the participants could select sites multiple times, the last site selection was used. *Fuel consumed* is measured by the change in fuel from the start and end of LPD. Since the AFM is steering the vehicle to the participant's intended location, fuel consumption is dependent on the final landing site, the time to complete, and the current state of the vehicle. *Landing site location* (m) are the coordinates relative to the baseline site (i.e., the center of the map). These three measures are collected without researcher intervention.

The last two measures were solicited during a debriefing session via a semi-structured interview. After the practice session, pilots were asked questions regarding their overall *strategy*. Strategy was defined as the course of action and associated weightings and preferences in determining where to land. Participants were asked to describe their decision making process, denote which scenario attributes were factored into the process, and the relative importance of each cue. *Workload* was self-reported using the Bedford Workload Scale [82] (Appendix A, Figure 42). Lower values denoted less workload. Both of these measures were collected after each run (six total per participant).

At the end of the experiment, participants were also asked to provide feedback on the displays. As human performance is dependent on the environmental context [83] [84], particularly as communication between humans and automation is through displays. Participants noted which display elements were used, with special emphasis on elements that factored into the decision making process, and suggested display improvements. Chapter 6 discusses the results of this analysis in greater detail.

Two dependent measures were calculated from the original set of data. *Landing Site Score* (LSS) is based on the results of Chua and Feigh [81]. A score is devised based on penalties each site receives with respect to certain site safety properties. A site can score from 0 to 1, with smaller scores denoting safer sites. The derivation of LSS is described in Equation 4. The participants were told that the vehicle was intending to land in the center of the map. An exact reason why the vehicle was landing there was not given. Therefore, the proximity to point of interest is defined as the Euclidean distance from the selected site and the center point of the map. Participants were asked to clearly state any assumptions they made regarding the vehicle operational performance and available utilities prior to and after touchdown.

3.4.1 Apparatus and Software

3.4.1.1 Hardware

The experiment was conducted in the Reconfigurable Operational Cockpit (ROC) at NASA Johnson Space Center in Houston, TX. To simulate the perspective of a true expansive environment, the ROC projects images from eight projectors on the side of its domed walls. The center of this dome holds a mock Orion vehicle cockpit, complete with three displays and one joystick for human input (Figure 9). The three displays can be individually set to present displays of any size and dimension, allowing for multiple configurations.

Two software programs, both run on Linux, were used for vehicle performance, display representation, and scenario simulation. The ALHAT Simulation, written in Trick [85], simulates vehicle performance and parts of the scenario. The ALHAT Simulation contains various elements including the AFM, the sensors, and Guidance, Navigation, and Control

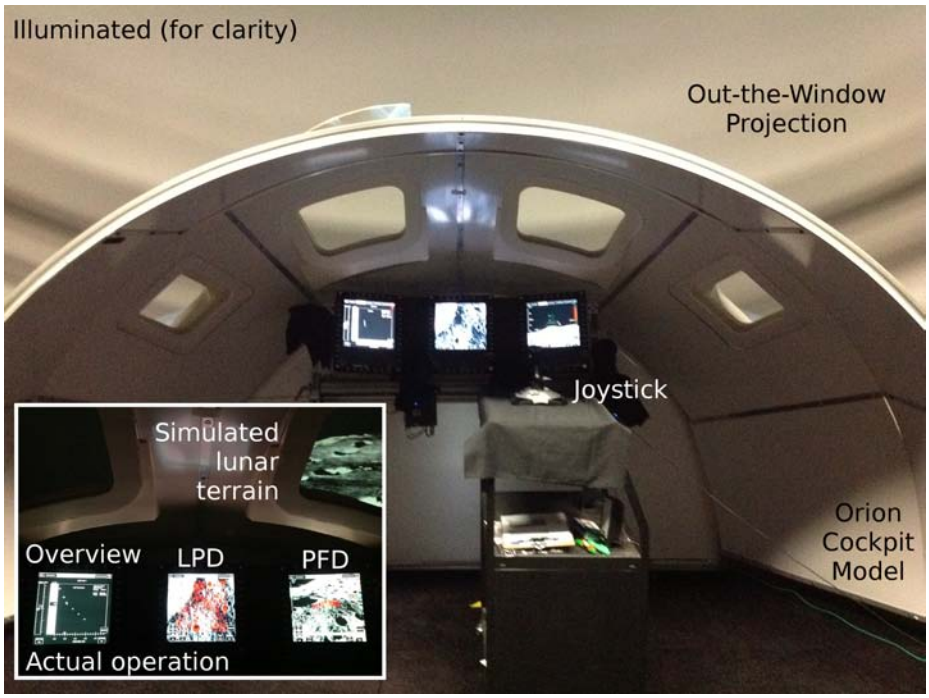


Figure 9: **ROC configuration.**

algorithms. Chapter 5 discusses this algorithm in greater detail, including a description of the reference vehicle.

NASA's Engineering DOUG [4] Graphics for Exploration (EDGE) software was used to simulate the displays and the lunar terrain perspective as seen from the cockpit. EDGE is written in Tool Command Language (tcl). The displays are modified from the designs developed by the ALHAT team [87, 88]. Each of the three displays is equipped with a progress bar; two clocks, Phase Time remaining and Time To Landing (TTL); Fuel remaining, as a mass and time measure; and Range, as a distance and time measurement.

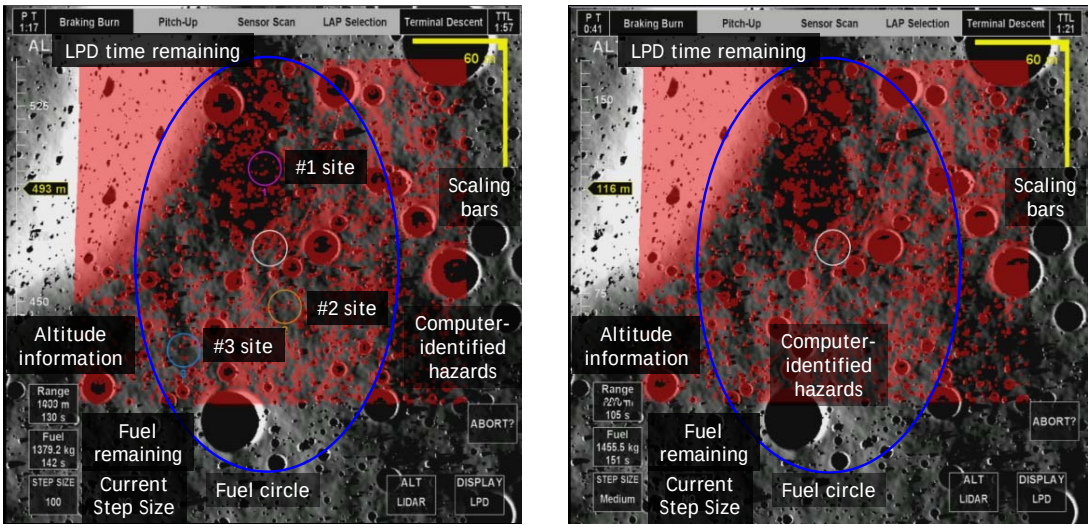
The overview display (Appendix A, Figure 48), located on the far left (Figure 9), presents a profile view of the vehicle and the trajectory through a ground range - altitude plot. Figure 48 illustrates a grayscale schematic of the diagram. The actual display is green lettering and lining on black. Breadcrumbs (filled squares) are "dropped" every five seconds to denote the previous state; predictors (empty squares) mark the projected state at five second intervals. This display also provides other state information, such as the altitude,

[4]The Dynamic Onboard Ubiquitous Graphics (DOUG) software is a NASA software product used for modeling, analysis, and training of International Space Station activity [86].

range, forward, and lateral velocities. All measurements on this display are listed in English units, based on the feedback of the last ALHAT crew evaluation [89].

The LPD display, located in the center, presents LPD-specific information on a top-down perspective of the expected landing area. This perspective features a photograph of the expected landing area and contains several overlays. The semi-transparent hazard identification overlay marks all hazardous craters, rocks, and slopes, in red. The fuel remaining overlay projects a blue ellipse. The area encircled by the ellipse represents the obtainable landing area based on the remaining fuel. The elliptical area shrinks in size as fuel is expended. The landing site overlay displays either three landing site options and a white circular marker or one landing site, depending on the automation role. In the high automation role, three numbered and colored sites are presented to the crew (Figure 10a). A white circular marker encircles a site when the corresponding button is selected. Each circle is equivalent to the lander footprint. When the "ENGAGE" button is pressed, the site options are removed, signaling the end of the participant's role in LPD. In the low automation role, a white landing site is provided (Figure 10b). Using the joystick, participants can move this site until they find a feasible spot. The flight computer follows this cursor around, updating the targeted location each instance the landing site changes. Participants can adjust the step size (large - 1000 in; medium - 500 in; small - 100 in) of the cursor by using the joystick buttons. This display also provides altitude information and a scale.

The PFD shares much of the same information as the LPD, but from a perspective. This viewpoint represents the feedback from a camera mounted on the front of the vehicle, angled at the intended landing site. The hazard identification, fuel remaining, and landing site overlays are all included, in addition to the altitude information. The PFD also includes a pitch ladder and a vertical speed indicator. Figure 47 is located in Appendix A, but does not include multi-toned shading as represented in the figure (stylistic change for presentation to the participants during the debriefing session).



(a) High Automation Role.

(b) Apollo Automation Role.

Figure 10: **Landing Point Designation Display.**

3.4.1.2 Scenario Generation

The lunar terrain was simulated using the Johns Hopkins University Applied Physics Laboratory's Digital Elevation Map maker (DEMmaker) software. DEMmaker was used to create the lunar maps and the associated shading. The DEMmaker software takes several inputs: the latitude and longitude of a central location anywhere on the Moon; a resolution (m/px); an $x-$ and $y-$ distance (km) or conversely, width and height (px). With this information, DEMmaker creates a DEM. The software refers to an internal database of major craters from the Goldstone Lunar Data and the Clementine mission and then uses mathematical models to random populate a distribution of smaller rocks and craters. The resultant DEM can be opened in APLNav for 3-D viewing (with correct latitude and longitude placement on a Moon grid) with lighting and slope analyses applied to the map. The user supplies a date-time string (*e.g.*, 2011-288T12-00-00.000) and the map will be shaded according to the celestial geometry and elevation of the mapped terrain. Similarly, the program calculates the changes in slope and colorizes the map based on the slope degree. This map making suite was used to generate 35 landing sites. The landing sites were shaded and analyzed regarding their slope profile. Candidate maps were eliminated based on several round of criteria. First, all maps with less than 30% or greater than 50% hazardous coverage were

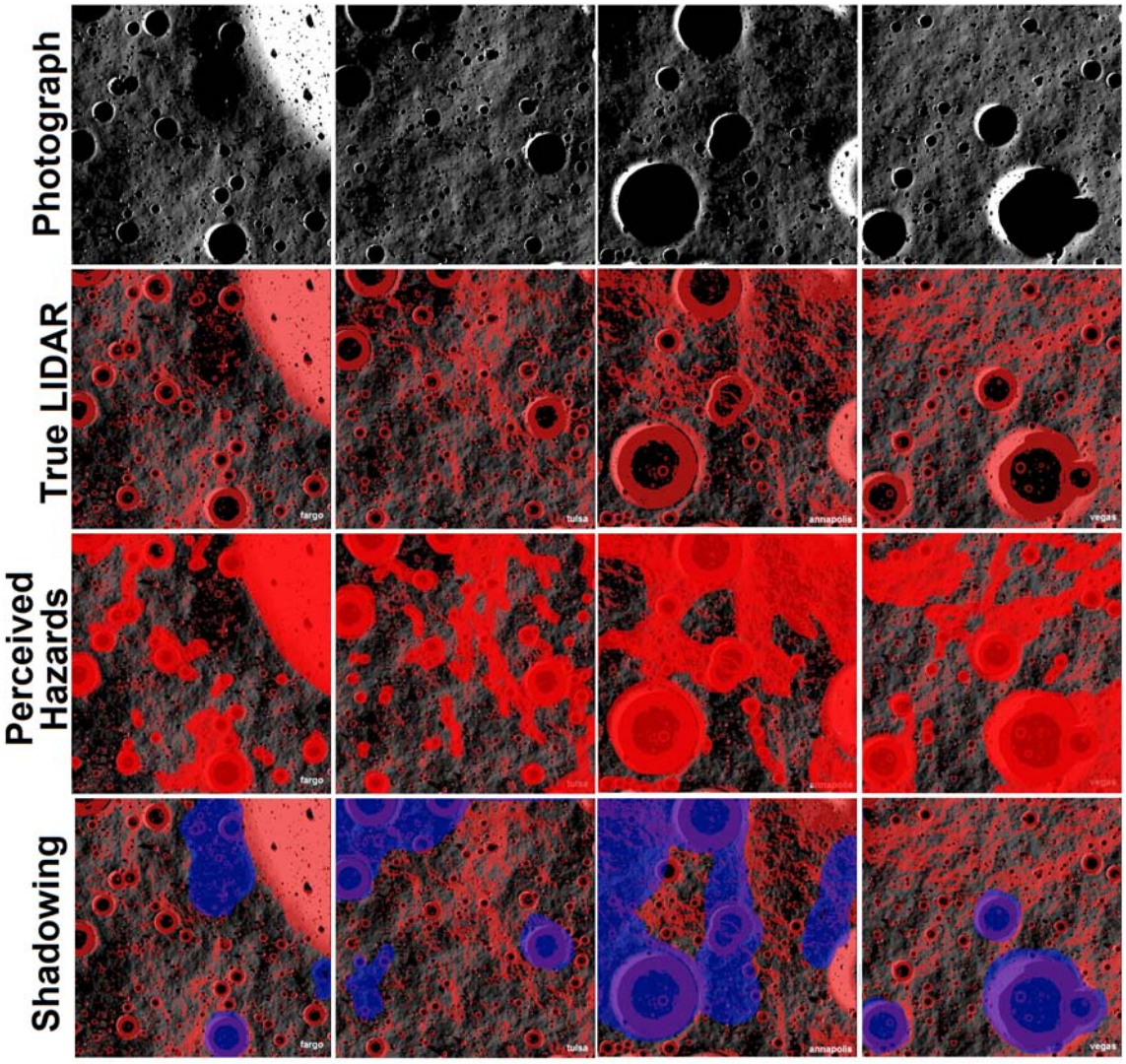


Figure 11: **Approximate perception of terrain by participants.**

eliminated. The remaining maps were evaluated based on the type of hazards at the point of interest (center of map), the shadowing of the photograph, and the perception of hazards. The shadow criterion was included based on participant comments from the study conducted by Chua et al. [90]. The last criterion is based on Gestalt principles [26], stating that disparate items placed closely together are perceived as one composite item. Maps were arranged based on quantity (hazard/no hazard; more/less shadowing; more/less perceived hazards) and similar maps were culled from this final group. The latitude and longitude of the four maps used in this experiment are listed in Table 12. All maps are 180×180m with a resolution of 0.2m per pixel. Figure 11 illustrates the properties of these maps.

The landing sites for the high automation maps were selected using the terrain sensing and recognition algorithm developed by NASA Jet Propulsion Laboratory [91] [92]. This algorithm uses LIDAR data to determine the probability of success in the landing area. The first three sites reported using this algorithm were used in this map for the nominal case. Since the LIDAR malfunction is based on erroneous data, sites were manually selected using the TSAR results as a guide.

The slope analysis provided by the APLNav software does not include any noise associated with sensor performance. Additional hazard scans were developed to account for changes in sensor performance due to two landing trajectories and two off-nominal scenarios. A LIDAR sensor cannot "see" past obstacles, meaning sensor noise is a function of position relative to the scan area. An algorithm was developed to determine the occlusion effects due to the baseline or shallow trajectory. Figure 12 describes the relationship between length of a shadow and position of the sensor. The premise of this algorithm is determining whether objects behind an arbitrary spot are in shadow. The viewing angle of point i, θ on the landing map is calculated using the arctangent of the LIDAR sensor position (y) relative to point i. The viewing angle, along with the altitude of point i, y_{shadow}, is used to determine the length of the shadow projected along the ground, x_{shadow}. The altitude of each point $i+1$ behind point i is checked to determine whether it is less than the expected shadow altitude. If the altitude of point $i+1$ is greater than the expected shadow altitude, the algorithm begins anew with this point.

As evidenced in Figure 13, the LIDAR sensor scan of the terrain is more obscured in the shallow than the baseline trajectory. The LIDAR malfunction scenario required the development of an extra hazard identification map. The LIDAR malfunction scenario was developed by overlaying two hazard scans, one associated with the location and the other from a different location. Both scans are truth scans; no performance effects due to trajectory were added. The composite hazard scan did not miss hazards, only adding falsely identified hazards. Participants were told however, that the LIDAR could have missed craters or slopes. The purpose for leaving true data rather than providing a completely erroneous scan was to prevent a complete disregard of the LIDAR scan. Participants were

Table 12: **Landing site locations.**

Scenario	**Lat, Long**[5]	**El., m**	**Site #1, m**	**Site #2, m**	**Site #3, m**
Baseline, nominal	-89.74, -240.02	770.5	-36.4, 2.3	26.7, -6.9	46.6, 39.5
Baseline, LIDAR	-89.74, -254.55	629.7	-34.5, 23.5	-47.3, -57.9	63.5, 72.5
Shallow, nominal	-89.66, -248.74	851.4	37.3, 51.5	-18.3, 45.9	-33.3, -23.9
Shallow, LIDAR	-89.70, -251.26	662.5	-30.5, -39.9	24.5, 56.2	-23.9, 35.5

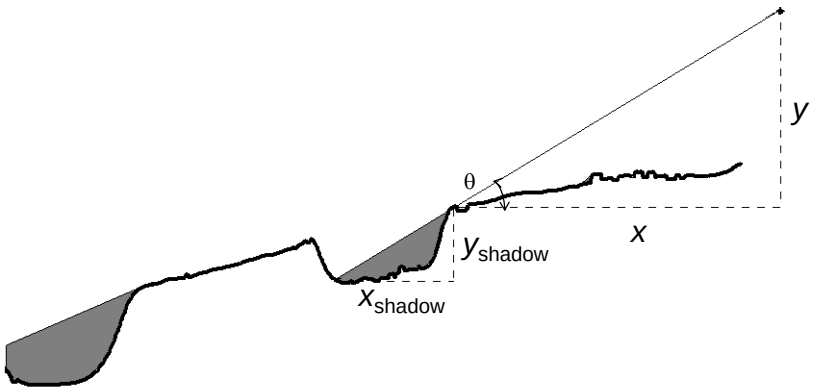


Figure 12: **LIDAR sensor position and occlusion effects.**

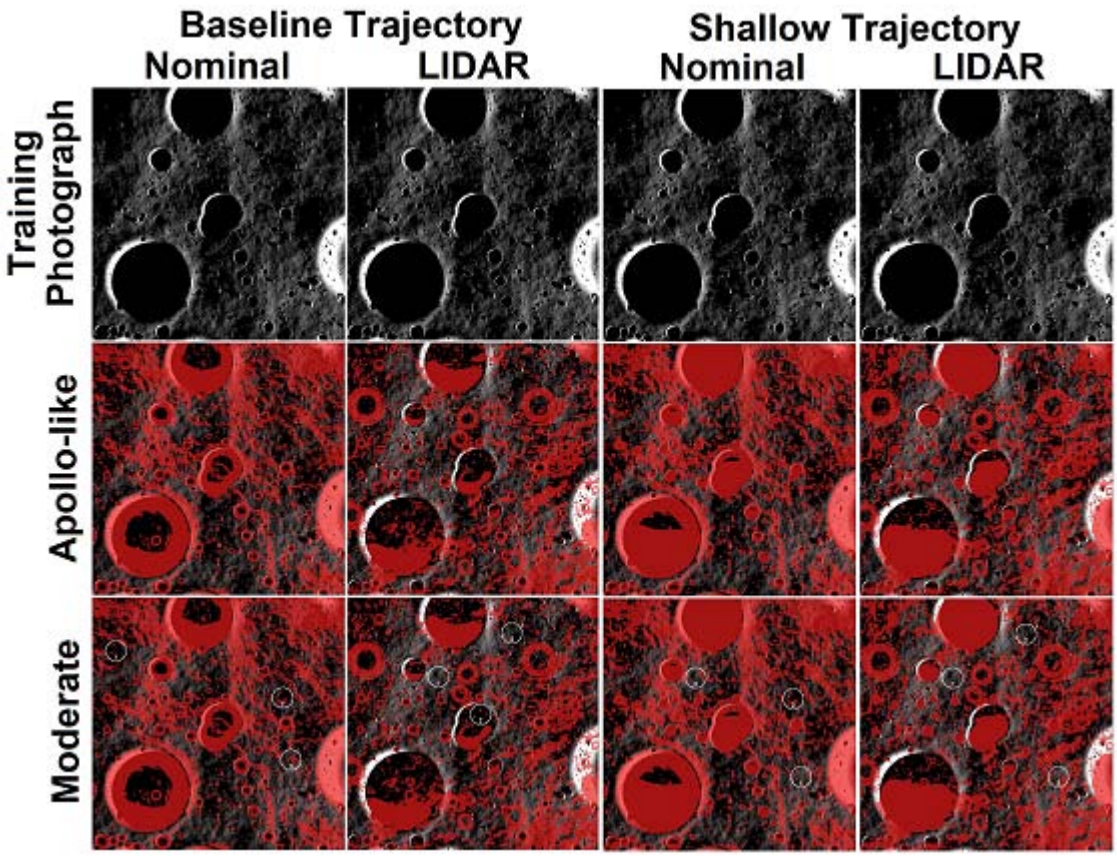


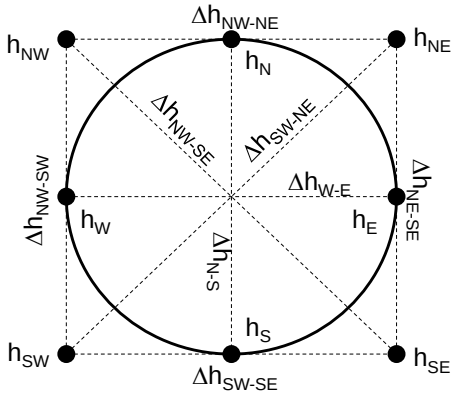
Figure 13: **Hazard scan and landing area photograph changes.** Four different landing areas were used in the actual experiment, but this example illustrates the differences due to the trajectory and the scenario. The baseline and shallow trajectories cause differences in LIDAR hazard identification. The LIDAR warning scenario identifies the major craters but also misidentifies hazards by excluding and including terrain attributes. The most distinctive mistake exists in the upper right and left hand corners in the LIDAR rows: two craters appear where there are none.

alerted to this sensor malfunction with a warning at the start of LPD: "LIDAR Sensor Warning". An example of the different maps and hazard scans produced for each landing area is presented in Figure 13.

In addition to generating realistic photographs and scans of all the landing areas, feedback materials were needed to prepare the participants for the experiment itself. It is difficult to train astronauts on a task that is still in infancy. Furthermore, the purpose of this experiment was to improve understanding of the astronaut decision making process during this task. Therefore, a delicate balance was necessary to provide sufficient feedback

to the participants without biasing their overall performance. For this experiment, participants received feedback on only the safety characteristics of the site. At the end of each practice run, participants were shown a contour map (gradient between green, yellow, and red - best to worst) of the safe regions. They were explicitly told this map represented "partial" feedback, and not meant to represent the criteria that their performance would be judged against. However, the same algorithm used to produce these contour maps was also used to define the Landing Site Score (LSS).

LSS is a function of three safety features: the induced slope due to roughness within the lander footprint, the proximity to hazards outside of the footprint, and the amount of hazards within the footprint. Induced slope is approximately calculated based on the largest slope produced by each of the lander pads. Figure 14 illustrates this geometry. Footprints with slopes greater than or equal to 10° are given a score of 1, the maximum penalty. All other footprints are assigned a score of 0 to 1, with zero penalty assigned to an ideal footprint of no slope. The proximity to hazards outside of the footprint is defined as the distance to the closest hazard. Footprints achieve no penalty if the closest hazard is at least one footprint, or one vehicle diameter, away. The amount of hazards within the footprint is calculated by determining the percentage of hazards within the footprint. Footprints are penalized corresponding to the percentage of hazards. Each of these attributes were weighted equally. Equations 4 summarize these penalty relationships.



$$S = \tan^{-1}\frac{\Delta h_i}{D_{\text{vehicle}}}$$

$$S_{\text{diagonal}} = \tan^{-1}\frac{\Delta h_i}{\sqrt{2}D_{\text{vehicle}}}$$

Figure 14: **Calculation of induced slope, or roughness, on lander due to terrain.**

$$\text{Induced Slope} = \begin{cases} \frac{S}{10} & S \leq 10^\circ \\ 1 & S > 10^\circ \end{cases} \tag{1}$$

$$\text{Proximity to Hazard} = \begin{cases} 1 - \frac{\min D_H}{D_{\text{vehicle}}} & \min D_H \leq D_{\text{vehicle}} \\ 0 & \min D_H > D_{\text{vehicle}} \end{cases} \tag{2}$$

$$\text{Percentage of Hazards} = A_{\text{Hazards}}/A_{\text{total}} \tag{3}$$

$$\text{LSS} = \frac{\text{Induced Slope + Proximity to Hazard + Percentage of Hazards}}{3} \text{LSS} \in [0, 1] \tag{4}$$

This experiment design was implemented over the course of three weeks in August 2012. The results and data analysis are discussed in the following Chapter.

CHAPTER IV

CHANGES IN ASTRONAUT DECISION MAKING

The creation of astronaut decision making models required understanding the strategies performed by astronauts during the Landing Point Designation (LPD) task, in two types of function allocation, trajectory, and operational scenarios. A strategy is defined as the search method, the combination of cues, and their relative importance used to reach a final conclusion. The study described in Chapter 3 discussed the design of a human-in-the-loop (HITL) experiment conducted to satisfy this main objective. Both qualitative and quantitative data were collected continuously through the experiment and during debriefing sessions conducted after each experiment run. The results of the qualitative analysis is presented first, followed by the statistical analysis of overall system performance. Lastly, this section concludes with discussion of the experiment design and a summary of the major findings. Chapter 5 discusses the incorporation of the experimental data within the cognitive process model.

Fourteen participants were involved in this study. As of the experiment start (July 2012), 80 former and current astronauts [93, 94] were active in the NASA Astronaut Corps. Ten participants were current members of the Corps, representing approximately $1/6^{\text{th}}$ of the overall targeted population. Two participants were experienced military helicopter pilots, and one participant was a recreational fixed wing pilot. All participants were employed in NASA Astronaut Office in administrative roles, as design engineers, or as cockpit designers. Only twelve participants provided data for the analysis presented in this chapter.

4.1 Qualitative Analysis and Results

4.1.1 Analysis of Observed strategies

The participants' decision making strategies were solicited through the debriefing questions, which were designed based on the previous work of Chua [78]. The individual answers were examined for similarity and combined in an attempt to demonstrate a universal strategy

Table 13: **Distribution of Participants Across Function Allocation and Decision Making Search Methods.**

Automation	Search Method	Participants	Total
Apollo-like	Local	105, 107, 112	3
	Areal	101, 108, 110	3
Moderate	Reranking	102, 106, 111, 113	4
	Eliminating	104, 109, 115	3

with several adaptations, accounting for the observed non-linear task order, iteration, and differences in task emphasis by individual pilots. As seen in Figure 15, all participants shared the centralized tasks of receiving data, interpreting data, searching and selecting an option. However, the search method varied depending on the automated assistance. Those using the automation system within the Apollo-like function allocation (henceforth shortened to "Apollo-like automation") reported using the areal or local search method. The *areal* method is conducted in two rounds. The first round consists of finding "good" regions where candidate sites may exist. In the second round, the participant then selects a landing site only from these regions. The *local* search method is focused on evaluating the entire landing area and finding a landing site. When using the automation system within the moderate function allocation (i.e., "moderate automation"), participants completed the task using either reranking or eliminating. The *reranking* search method evaluates all three candidate sites simultaneously and selects the best site of the three. The *eliminating* method is focused on eliminating sites until one site remains. Generally, only one elimination round is necessary. These definitions were self-assigned by participants, answering the debriefing question (listed in Appendix A). Table 13 illustrates this distribution of automation, search method, and participant.

Another common subtask completed by participants of both function allocations is the development of a mental model. One type of a mental model of the LPD task consists of determining where "good" and "bad" regions were located. Others may be models of the vehicle state and the LPD sequence. However, the focus is strictly on anticipated choice of landing areas. This exercise was completed by half of the participants with a photograph of the expected landing area prior to each experiment run. Most participants defined good

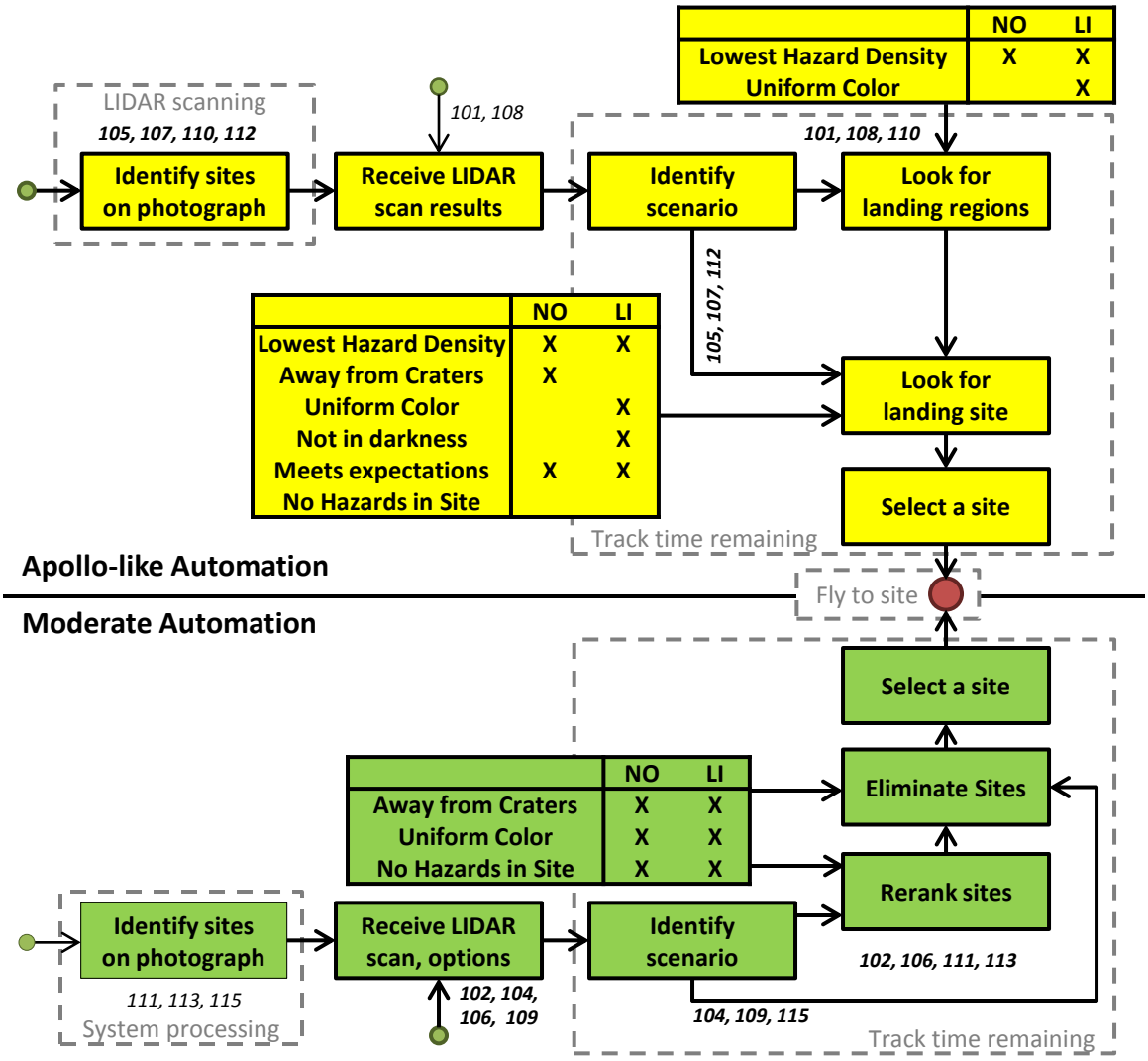
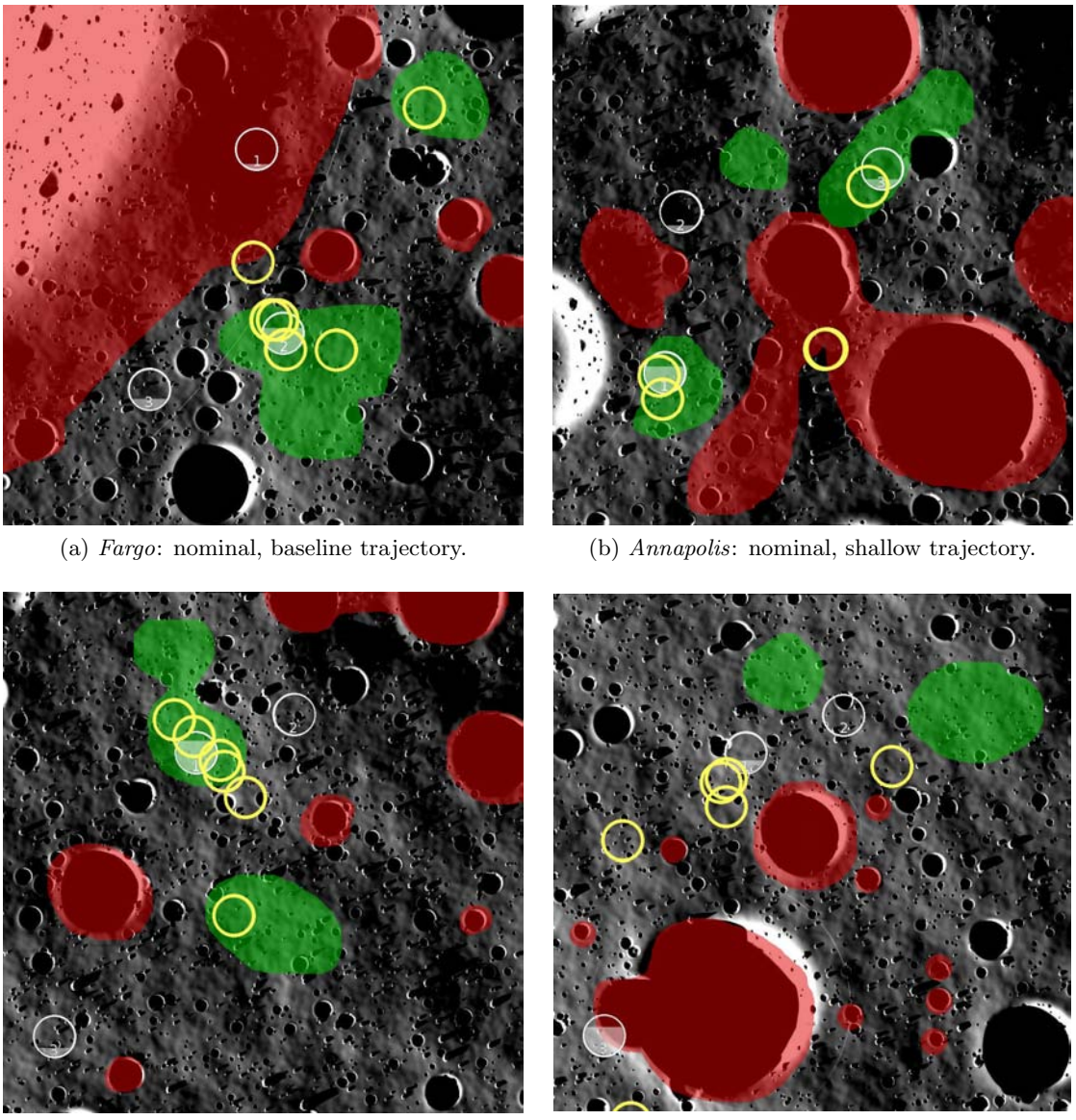


Figure 15: **Landing Point Designation Strategy for Apollo-like and Moderate Function Allocations.** The chart denotes the start of the LPD task with the leftmost dots - there are two entry points depending on whether the participant used the photograph of the expected landing area. The task ends at the rightmost dot. Shaded boxes in bold outline are subtasks used by at least half of the participants within each automation sequence. This condition is defined as at least three participants in the Apollo automation and at least four participants in the moderate automation. Each automation is colored differently - yellow for Apollo and green for moderate. Boxes in white and in thin outline are subtasks used by a small portion of the participants. Participant identifiers ($10N$ number) along the task sequence are those who demonstrated that particular subtask. Dashed lines indicate simulation/automation activity concurrent with participant activity.

regions as those with relatively uniform shading distribution (interpreted as low slope), away from major craters and large shading gradients (interpreted as significant altitude change). Bad areas are essentially hazards or unknowns: craters, large shading gradients, and dark regions. After each run, participants who stated that they had formulated this mental map were asked to mark these regions on the respective photograph. The annotations provided by the fourteen participants were superimposed on each other to form Figures 16a-16d, which illustrates the combined opinions. The location of the three automation-suggested sites are also shown in these figures, but were not initially shown to the participants until the simulation began. Plots of participant selections with respect to the combined mental models are listed in Figure 16. Plots with respect to the printed LIDAR scan are shown in Appendix A, Figure 49. Generally, participants selected sites near good areas. This development of the mental model is analogous to the action of "gating" as performed by the Apollo astronauts. During high and low gates, the Commander and Pilot were expected to make checks associated with the vehicle state, health, and mission status. Knowing when to establish these checks, or gates, is an aspect related to training that must be carefully decided.

Participants using the Apollo-like automation had the option of refining, or adjusting, their initially selected landing site. Since the start of refinement was not directly measured (the participant did not indicate entering a refinement mode), this time duration was inferred from timestamped joystick maneuvers. The start of refinement was characterized by a reduction of the step size used to maneuver the designated landing site icon on the display. However, minimizing the step size was not always a clear indicator of refinement, as the minimum step size was used through the entire run. The second indicator was the path of the designated landing site icon. Generally, if the participant backtracked more than one step or circled around the initial landing site, he was performing refinement. The difference in timestamps was equivalent to the refine process. An example of these two refinement start conditions is presented in Figure 17. The left movement pattern represents a participant who moved the designated site icon upwards on his display, then left two units, then downwards, and circled a central area that contained his final landing site. The downward



(a) *Fargo*: nominal, baseline trajectory.

(b) *Annapolis*: nominal, shallow trajectory.

(c) *Tulsa*: LIDAR warning, baseline trajectory.

(d) *Vegas*: LIDAR warning, shallow trajectory.

Figure 16: **Representation of the astronaut mental model.** Red spots are participant-identified bad areas; green are good areas. The three white landing sites ranked by the moderate automation are shaded according to the percentage of participants who selected this site. The yellow landing sites are those selected by participants with the Apollo-like automation.

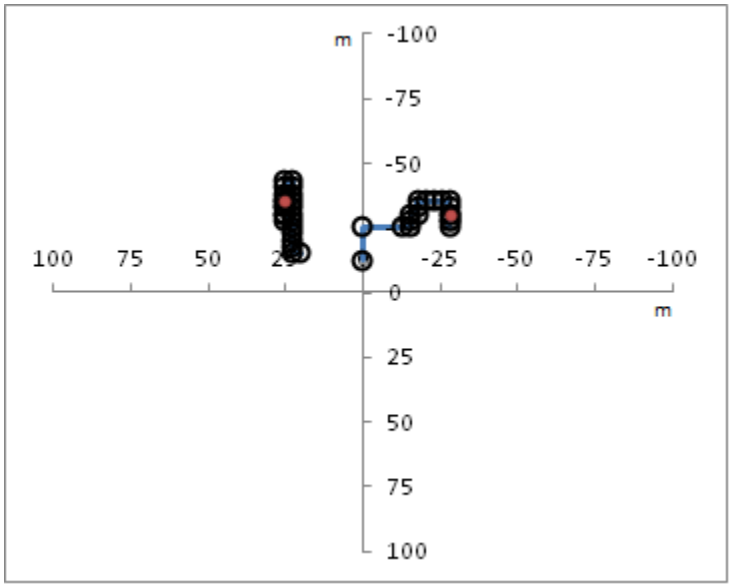


Figure 17: **Example of projected landing site movements during two separate experiment runs.** The final landing site chosen by participants is marked by a red dot and the time stamped movements are in black circles. A blue line traces the path of the icon.

movement of the designated site icon is the refinement start. The right movement pattern is representative of two designations and no refinement. The first initial decision making period is indicated by a change in step size. However, he does contain himself to this area and uses this minimum step size to find another site. He is observed to overshoot his landing site by one unit and does not refine.

Figure 18 illustrates the distribution of time spent during the initial decision making, the refinement, and the automation executing the final participant decision. This analysis was used to determine if more or less refinement time was induced by a scenario, a trajectory type, a search method, or background. No significant correlations were observed. The distribution of activity indicates that participants did not aim to select an initial landing site early and then refine the site as the vehicle touched down, as hypothesized. However, if the participants were maneuvering the vehicle to the intended landing site, it is likely refine would occur through the entire process.

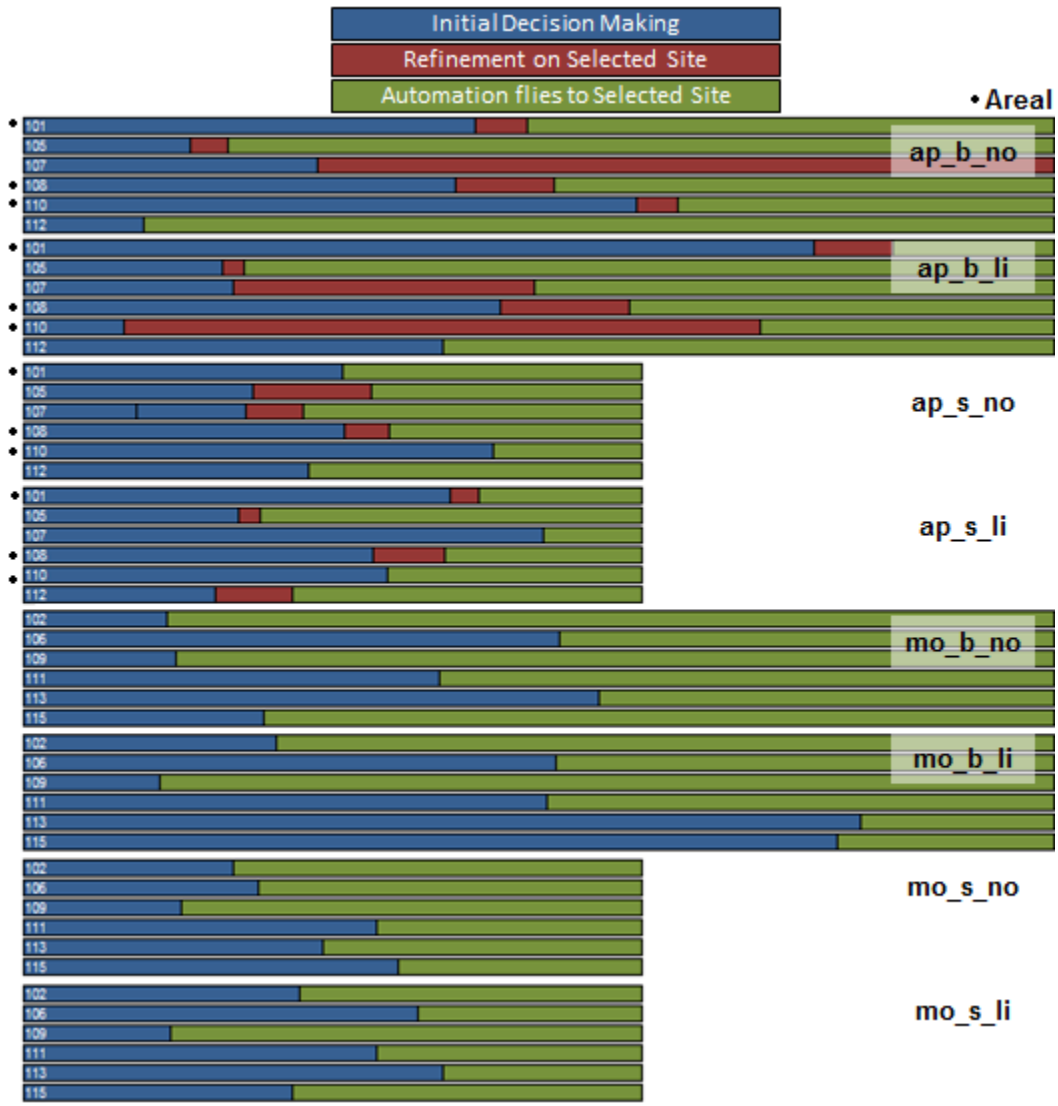


Figure 18: **Distribution of activities during landing point designation.** The naming convention is as follows: *Function Allocation* (ap - Apollo-like/mo - Moderate), *Trajectory* (b - Baseline/s - Shallow), *Scenario* (no - Nominal/li - LIDAR Warning).

4.1.2 Analysis of the Relative Importance of Decision Making Criteria

As seen in Figure 15, participants using different function allocations and search methods placed emphasis on different decision making cues. Based on previous experiments [78, 81], participants could use any combination of sixteen different cues, ten of which can be further defined by quantity ("have/did not have"), six of which are threshold ("within/by"). Table 14 represents this full spectrum.

Participants reported using ten of these cues, which can be categorized into three general groups: system safety (no hazards within landing area, away from craters, lowest hazard density, uniform color, not in complete darkness, minimize fuel consumption, work faster), mission objectives (located near point of interest (POI)) and automation rank. An eleventh cue, located in the upper two-thirds of the display, was reported and is considered as part of system safety.

Participants were most concerned with attributes of system safety. Generally, participants preferred and actively sought sites that contained *no hazards within the landing area*. However, if no such site existed, participants focused on the distribution and location of hazards within the landing site. Almost all participants assumed that the lander feet would be at the vertices of a perfect square contained within the landing circle. The lander would not rotate around the central body axis. Subsequently, some participants preferred sites with hazards widely distributed, or not on the edge of the landing circle, or conversely, not in the center of the landing site. The possible collision of a central hazard with the main engine was of particular concern to a few participants. One participant commented on the lack of visible lander pads from either the display or the window. He believed that without this feedback, the LPD task could not be successfully completed. The final maneuvers prior to vehicle touchdown are minute adjustments to the vehicle position, as to avoid landing on any rocks or craters.

The structural design of the lander vehicle and the camera angles provided to the participants resulted in additional landing site criteria. The participant who required lander pad visibility stated that his landing performances were arbitrary - he was effectively guessing at where to land. Subsequently, his quantitative performance has not been included in this

Table 14: **List of possible cues for Landing Point Designation.** Overall, there were sixteen cues that could be used to determine where to land on the Moon. The left column indicates those cues used by participants during the experiment. The right column illustrates the category of cue and the possible inclination. Follow-up questions to each cue are noted in the footnotes.

Used?	Cue
	Sites that **didn't have/had** hazards along the approach [a].
x	Sites that are **away/near** from large craters [b].
x	Areas that have the **lowest/highest** hazard density.
x	Sites that **don't have/have** hazards inside the landing area [c].
x	Sites that **don't have/have** a uniform color with in the landing site [d] [e]
x	Sites that **aren't/are** in complete darkness.
x	Sites that are located **downrange/reverse downrange/center/left/right** of the point of interest.
x	Sites with a **higher/lower** automation ranking.
x	Sites within previously identified **good/bad** regions.
x	I performed LPD **lower/faster** than normal.
x	I attempted to perform LPD **within/by** ____ kg of fuel.
	I attempted to perform LPD **within/by** ____ m in altitude.
	I attempted to perform LPD **within/by** ____ deg of pitch.
	I attempted to perform LPD **within/by** ____ m/s in velocity.
	I attempted to perform LPD **within/by** ____ m in range.
x	Sites that are in the **upper/lower** two-thirds of the display.

[a] How far along the approach, in vehicle diameters?
[b] What is a large crater, in vehicle diameters?
[c] If placement matters, where should hazards not be?
[d] What color, and what is your interpretation of the color?
[e] Are you searching for a particular color pattern?

statistical analysis. Another participant was frustrated with the limitations in the camera angle. As the vehicle drew closer to the surface, the terrain immediately underneath the vehicle became obscured by the main frame in both the PFD and the window view. This lack of continuous visual information was factored in by the participant. He reported only selecting sites that were *located in the upper two-thirds of the display.*

Landing sites *away from craters* were also of interest to participants. Previous studies [78] illustrated that participants preferred landing sites that were at least a landing site diameter, D, away from large craters. In this study, some participants agreed with this definition (a buffer of one landing site diameter), but many participants were comfortable with landing near craters. At least two participants mentioned scientific value of landing near large craters. The definition of a large crater also differed between participants. Seven participants provided definitions on what constituted a large crater; large craters were defined as between 0.25 D to 1.5 D. About half of the participants focused on identifying good regions prior to selecting a specific landing site. These good regions were characterized by *low hazard density.* Low hazard density regions typically do not have many large hazards and minor hazards are widely distributed.

The shading of the maps revealed a great deal of information regarding the terrain. This information was available as a background overlay with the hazard identification, out the window, and with the photograph of the expected landing area presented prior to each run. Participants were particularly concerned with the slope of the terrain, both within the landing site and immediate outside. Since slope information was not explicitly given to the pilots (in terms of a degree or a contour line), the change in pixel shading was used as an indicator. In general, participants preferred sites with a *uniform shade* within the landing site. Since the sun is at an angle relative to the surface, different altitudes will be lit to differing degrees. A uniform shade denoted flat terrain, or low gradients. The shading tone itself was critical: participants preferred sites of a lighter shade (interpreted as higher in altitude), or gray to white, as it was believed to have greater visibility. Participants also searched for sites that were *not in complete darkness.* Dark landing sites were not appealing as the participant could not assess the terrain; these sites were typically eliminated from

consideration. Participants were not informed, nor did any ask, if the lunar lander was equipped with external lighting capabilities. Half of the participants ignored sites that were in complete darkness.

Participants were also concerned with *fuel consumption*. The Guidance, Navigation, and Control (GNC) algorithms within the simulation were robust enough within the 180 × 180 m landing area. The participant, unbeknownst to them, were not threatened by lack of fuel at any point within the landing sequence. Despite having this information available to them through the entire LPD sequence, no participant indicated a sense of the total kilograms of fuel consumed during the LPD task. However, several participants still factored in fuel consumption. They attempted to make short diverts from the center point, or landing close to the center of the map. Additionally, several participants reported trying to *work faster*, as to save fuel. Some participants opted to land short (bottom half of the map), especially during off-nominal situations. Their reasoning was to land quickly and avoid any further complications with the vehicle. Participants felt rushed more in the off-nominal scenarios.

Unsurprisingly, all participants were concerned with system safety and no participant chose a landing site strictly on mission objectives. Criteria use differed based on the function allocation and scenario. Table 15 illustrates criteria used by the number of participants. Different scenarios resulted in different cue usage. In the nominal scenario, the Apollo-like pilots were particularly concerned with finding sites in areas with fewer hazards that matched training expectations. This behavior is expected of well-trained, experienced astronauts. In contrast, the LIDAR warning scenario shifted reliance on expectations to finding a site that is in an area free of hazards and of uniform color. The participants were accounting for faulty LIDAR data by avoiding all identified hazards and uneven terrain. The difference in participant role between the Apollo-like and moderate function allocation is reflected in the cue usage. Since the moderate mode provides only three choices that are clearly marked on the display, the participants focused more on the sites themselves, rather than areas. The difference is most prominent as they sought hazard-free sites, rather than areas of lower hazard density. These same preferences remained constant across both types.

In general, participants used between one and five cues to decide where to land, with

Table 15: **Distribution of Cue Usage per Scenario.**

Apollo-Like			**Moderate**		
	Nom.	LI		Nom	LI
Lowest hazard density	3	3	No hazards in landing	7	7
No hazards in landing	3	3	Away from craters	5	6
Meets expectations	3	2	Uniform coloring	4	5
Uniform color	2	3	Located near POI	2	2
Located near POI	2	1	Not in darkness	1	3
Not in darkness	2	2	Meets expectations	2	2
Located upper 2/3	1	1	Lowest hazard density	2	1
Away from craters	-	1			
Automation Rank	N/A	N/A	Automation Rank	1	1

an average of three cues for every scenario. Relative importance, or ranking, rather than weightings, were assigned to each cue. Patterns in cue usage are listed in Table 16. A comprehensive summary of each participant's cues is listed in Appendix A, Table 27.

4.2 Quantitative Analysis and Results

This experiment was designed to determine whether automation, trajectory, and scenario had a significant effect on overall system performance. All tests were conducted at $\alpha = 0.1$ to determine significant trends as the number of participants was particularly low. Unless otherwise noted, a Mixed Measures ANOVA was used for the analysis.

4.2.1 Overall statistical results

Overall, participants completed the LPD task in $\mu_t = 28.66$s ($\sigma_t = 15.15$), with a range of 8.63 to 74.95s. During this task, they were able to achieve an average Landing Site Score (LSS) of $\mu_{\text{LSS}} = 0.5803$ ($\sigma_{\text{LSS}} = 0.132$). On average, the Proximity to Point of Interest (POI) score was $\mu_{\text{POI}} = 0.459$ ($\sigma_{\text{POI}} = 0.135$). Both POI scores and LSS were on a range between 0 and 1, with 0 equaling a perfect score and 1 being the worst score. Participants used an average of $\mu_{\text{fuel}} = 544.77$ kg ($\sigma_{\text{fuel}} = 137.01$) on a fuel consumption range of 331.30 to 702.49 kg. Overall, participants perceived this task to be a mean workload of $\mu_{\text{work}} = 3.712$ ($\sigma_{\text{work}} = 1.576$). Varied opinions were expressed on the workload intensity of the task, with participants ranging from 1 ("Workload insignificant") to 9 ("Extremely high workload. No spare capacity. Serious doubts as to ability to maintain level of effort").

Table 16: **Cue usage patterns.**

Apollo-like automation, nominal
1. No hazards in landing is ranked highest
2. No hazards in landing and meets expectations are usually used together (baseline trajectory only)
3. No hazards in landing is equal to or more important than expectations (baseline trajectory only)
4. Located near POI, if used, is generally ranked low
5. No hazards in landing and lowest hazard density cannot be used in the same round of evaluation.
Moderate automation, nominal
1. No hazards in landing is always used.
2. No hazards in landing and Away from craters are ranked highest
3. No hazards in landing and away from craters are usually used together
4. No hazards in landing and lowest hazard density cannot be used in the same round of evaluation.
Apollo-like automation, LIDAR warning
1. No hazards in landing area and uniform coloring are usually ranked highest
Moderate automation, LIDAR warning
1. Away from craters is always used.
s No hazards in landing and away from craters are usually used together
2.
3. Away from craters and uniform color are usually used together (shallow trajectory only)

Figures 19 and 20 illustrate the distribution of these values.

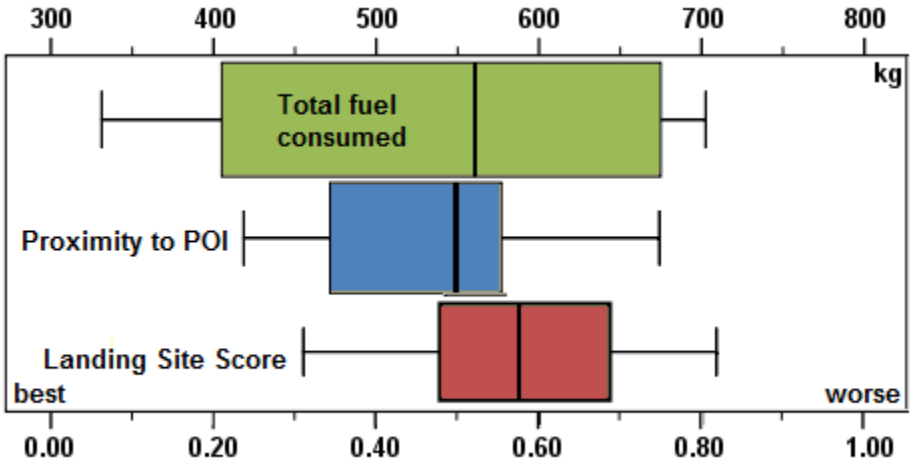


Figure 19: **Distribution of participant performance scores with respect to landing site selection, proximity to POI and fuel consumption.**

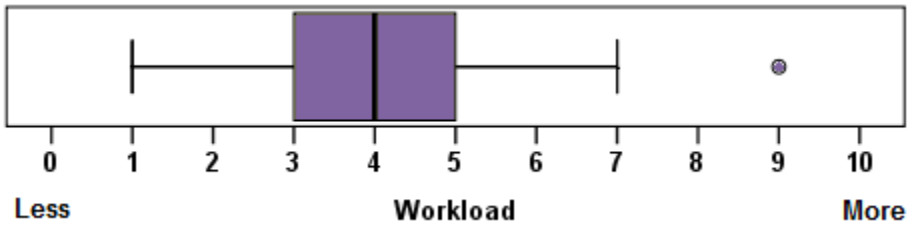


Figure 20: **Distribution of participant workload scores.**

4.2.2 Impact of Function Allocation, Trajectory, and Scenario on LPD Performance and Astronaut Workload

This experiment was designed to examine the effects of function allocation, trajectory, and malfunctioning LIDAR on overall LPD performance (task completion time, fuel consumption, LSS, POI) and astronaut workload. Results from the MANOVA illustrated that automation ($\Lambda = 0.549, p < 0.000$), trajectory ($\Lambda = 0.008, p < 0.000$), and scenario ($\Lambda = 0.344, p < 0.000$) all had an effect on overall performance (Figure 21). Mixed measures ANOVA was used as a follow up to the MANOVA, to determine the specific effect of each independent variable. Function allocation significantly affected LSS

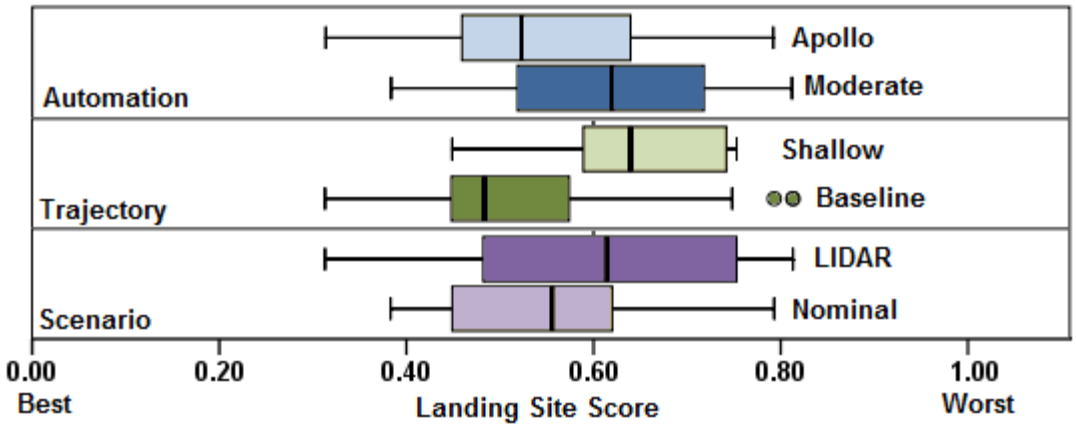


Figure 21: **Distribution of landing site score across function allocation, trajectory, and scenario type.**

($F(1,11) = 5.372, p < 0.041$) and fuel ($F(1,11) = 10.167, p < 0.009$). Participants generally chose safer sites when using the Apollo-like automation compared to moderate automation and also consumed less fuel (534.5 kg to 553.6 kg). Trajectory had a significant effect on the fuel consumption ($F(1,11) = 6224.606, p < 0.000$). The baseline trajectory consumed more fuel than the shallow trajectory. However, participants chose safer sites in the baseline ($F(1,11) = 19.189, p < 0.001, \mu_{\text{b}} = 0.519$) than in the shallow trajectory ($\mu_{\text{s}} = 0.642$). The type of trajectory also had a significant impact on the completion time. Participants performed the task faster in the shallow ($F(1,11) = 9.791, p < 0.010, \mu_{\text{s}} = 23.07\text{s}$) than in the baseline trajectory ($\mu_{\text{b}} = 34.25\text{s}$). Lastly, participants chose sites closer to the POI in the baseline ($F(1,11) = 5.344, p < 0.041, \mu_{\text{b}} = 0.420\text{s}$) compared to the shallow trajectory ($\mu_{\text{s}} = 0.4977\text{s}$). The type of scenario was also observed to have a significant effect on performance. Participants chose less safe sites in the LIDAR warning scenario than in the nominal ($F(1,11) = 105.567, p < 0.00$).

Interactions between the main effects were observed to be significant on overall performance and workload. The interaction between trajectory and scenario indicated that sites selected during a nominal baseline approach were the most safe ($F(2) = 4.101, p < 0.05, \mu_{\text{LSS-b-no}} = 0.440$) and as expected, the sites selected during the shallow approach during the LIDAR warning scenario were the least safe ($\mu_{\text{LSS-s-li}} = 0.695$). These trends are

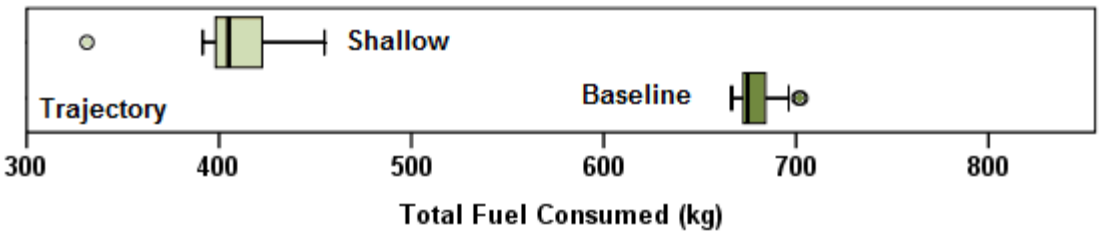


Figure 22: **Distribution of fuel consumption across trajectory types.**

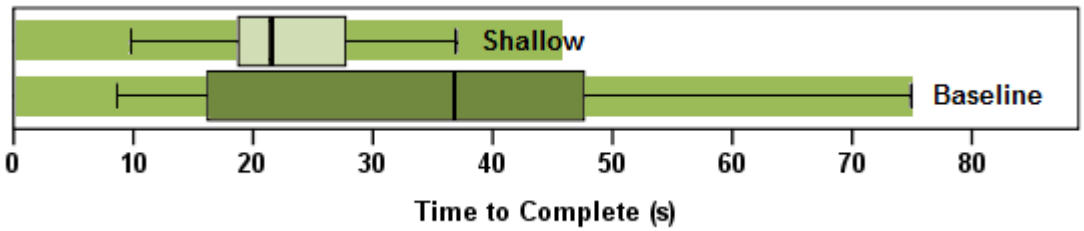


Figure 23: **Distribution of time to complete across trajectory types.**

seen in Figure 24. Similarly, sites were picked closer to the POI of interest due to trajectory and scenario interactions ($F(2) = 8.210, p < 0.005$). Sites selected on a baseline approach during the LIDAR warning scenario were closest to the center ($\mu_{\text{prox-b-li}} = 0.380$) whereas the shallow approach in the same scenario are farthest ($\mu_{\text{prox-s-li}} = 0.575$) (Figure 24). The interaction between trajectory and scenario and function allocation was marginally significant on the overall LPD performance.

Workload was examined separately from the other dependent variables as it was self reported. None of the individual main effects - function allocation, trajectory, or scenario -

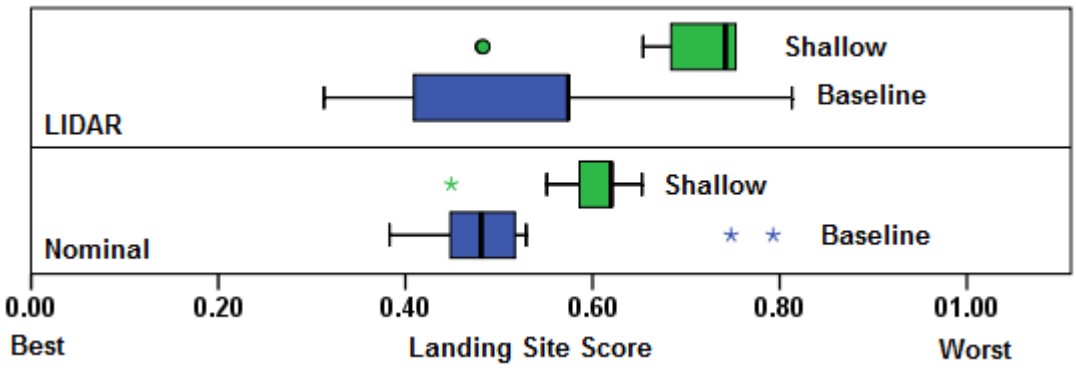


Figure 24: **Distribution of LSS due to trajectory-scenario interaction.**

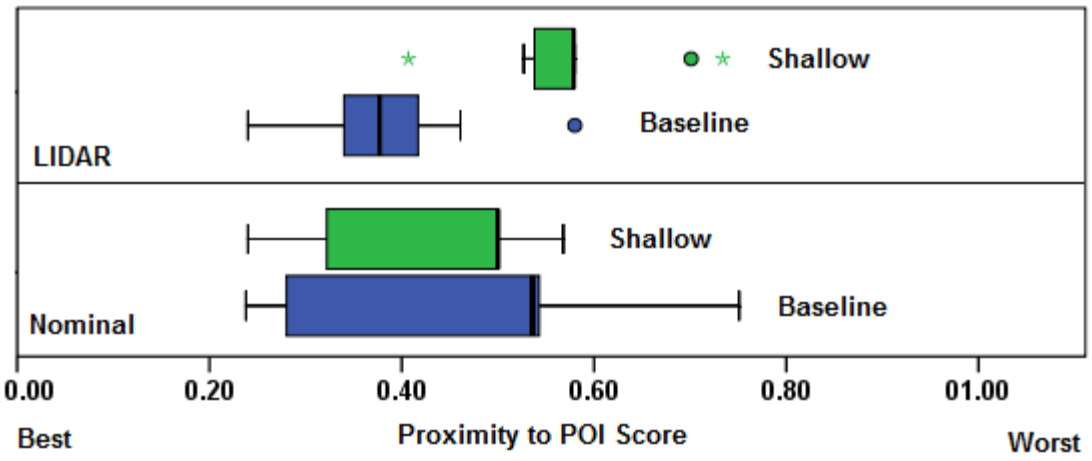


Figure 25: **Distribution of Proximity to POI score due to trajectory-scenario interaction.**

significantly impacted workload, with respective powers of 0.294, 0.056, and 0.138.

Pearson correlations were also performed to determine whether relationships between cue usage and task completion time exist, based on the results presented in Table 15. There was no significant correlation between number of cues and completion time, nor was there a significant correlation between the types of cues and completion time.

4.2.3 Impact of Participant Background on LPD Performance and Astronaut Workload

Since the participants had a variety of astronaut training, fixed-wing, rotorcraft, and spacecraft flight experience, it was hypothesized that certain demographics would perform the LPD task differently. Results from an ANCOVA showed that a participant's highest level of training was not significant on LPD performance. However, this level of training is significant on the self-reported workload score. Pilots and commanders rated the LPD task low ($\mu_{\text{pilots}} = 1.5$), whereas MS/Astronaut Candidacy graduates (defined as having completed Astronaut training but no spaceflight experience) and helicopter pilots ranked the task higher ($\mu_{\text{MS/AsCan}} = 3.85$ and $\mu_{\text{heli}} = 3.67$, respectively).

4.2.4 Impact of Variable Weightings on Performance

The statistical analysis was completed on individual dimensions of performance (fuel consumption, time to complete, LSS, POI), with a MANOVA test to examine the dependencies

between these dimensions. However, performance is often defined as a weighted combination of these dimensions. The relative importance of each criterion for future missions to the Moon has not been defined. Subsequently, it is of interest to mission designers to examine the changes in performance as criteria priorities are shifted. This particular analysis considered performance based on LSS and POI. The GNC algorithm employed in this experiment was robust to landing site selections within the expected landing area and trajectory profile, thus rendering fuel consumption and time to complete differences between groups negligible.

Figures 26, 27, 28, 29 and 30 were each plotted on a POI weighting - performance score plot. A value of 0.1 on the x-axis implies that the POI weighting is 0.1 and LSS, or safety, is at 0.9. Axes labels indicate greater emphasis on safety or proximity and whether performance is better or worse. The central thick line is the mean score of all points within the specified group, using that particular combination of safety and proximity to POI weights. The dashed lines are 1-σ from the mean. The thinner, spotted lines are the minimum scores within the data set for that particular group.

Figure 26 compares the overall performance of the automation system in a robotic function allocation and participants. The automation data are derived from the proximity to POI and safety score of the highest ranked site in the Moderate automation runs. It is the equivalent of a fully robotic case wherein the automation system operates the LIDAR and selects a site within 5s. There are four points overall, to reflect the robotic selection at a) nominal operations, baseline trajectory; b) nominal operations, shallow trajectory; c) off-nominal operations, baseline trajectory; d) off-nominal operations, shallow trajectory. The participants' data include all runs, across all of the independent variables. As illustrated in this graph, some participants were better at choosing sites closer to the POI than the robotic function allocation. In general, however, participants do not choose better sites than the robotic function automation, particularly when safety is the highest priority. The difference between means is less than one sigma in standard deviation.

Figure 27 examines the changes in performance as split between Apollo-like and moderate function allocation. The data are from the respective participants' performances during

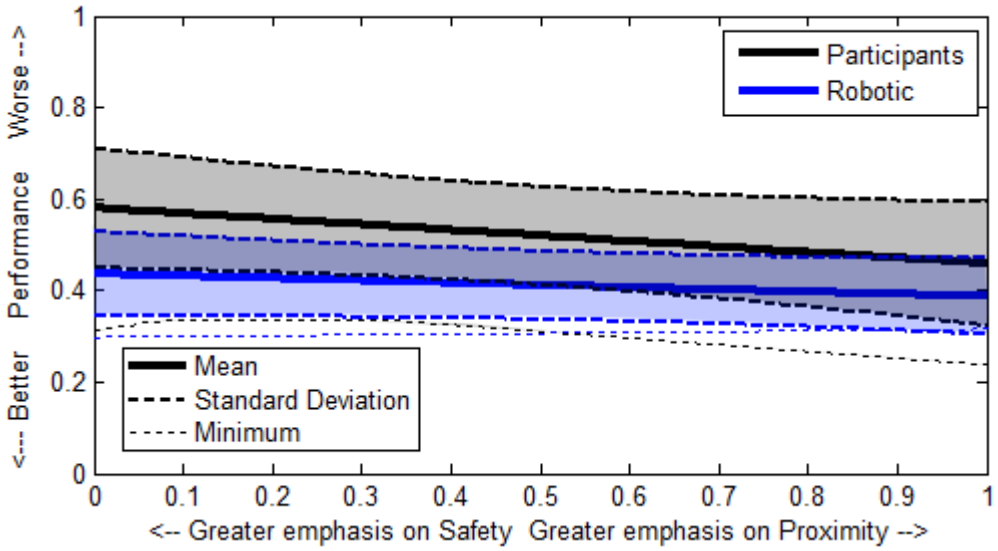


Figure 26: **Performance changes between participants and automation with variable weighting.**

the simulated LPD landing. There are 24 data points for Apollo-like and 28 points for moderate. As seen in this figure, both groups performed about equally as proximity becomes a greater priority. However, participants using the Apollo-like automation performed slightly better when safety was the primary measure of performance. The difference in the mean of the performance scores is not one-sigma or greater, and they are approximately equal as they are within 0.1 along the LSS [1].

Figure 28 evaluates the performance differences between the baseline and shallow trajectories. Each trajectory group includes data from both Apollo-like and moderate function allocation groups, resulting in an even split of 26 data points per trajectory. The performance during the baseline trajectory is better than the shallow trajectory, regardless of the priority between proximity and safety.

Figure 29 illustrates the performance differences between the two scenario types: nominal and LIDAR warning. Each scenario type includes data from both Apollo-like and moderate function allocation, in addition to both trajectories. There are 26 data points in

[1] see Chapter 5 for an explanation of this derivation

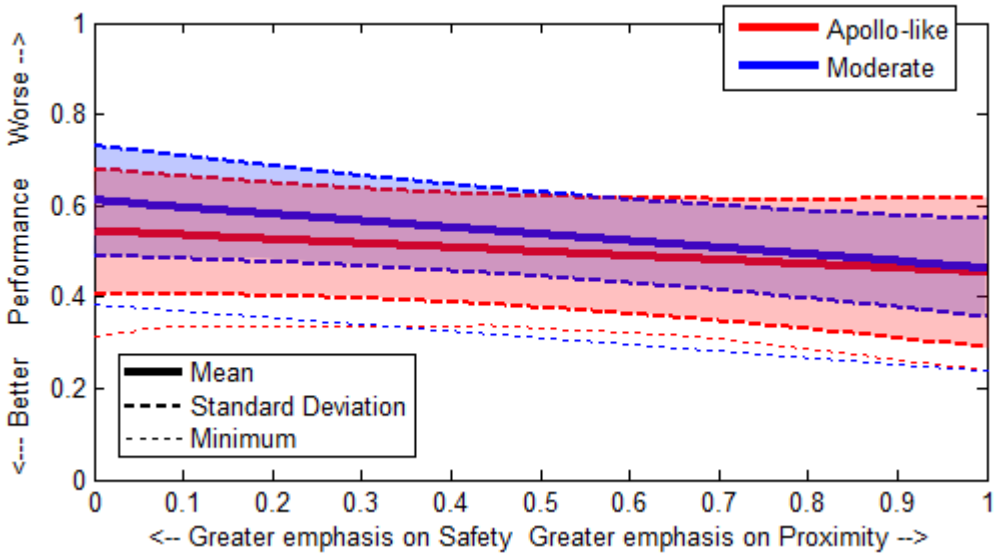


Figure 27: **Performance changes across function allocation with variable weighting.**

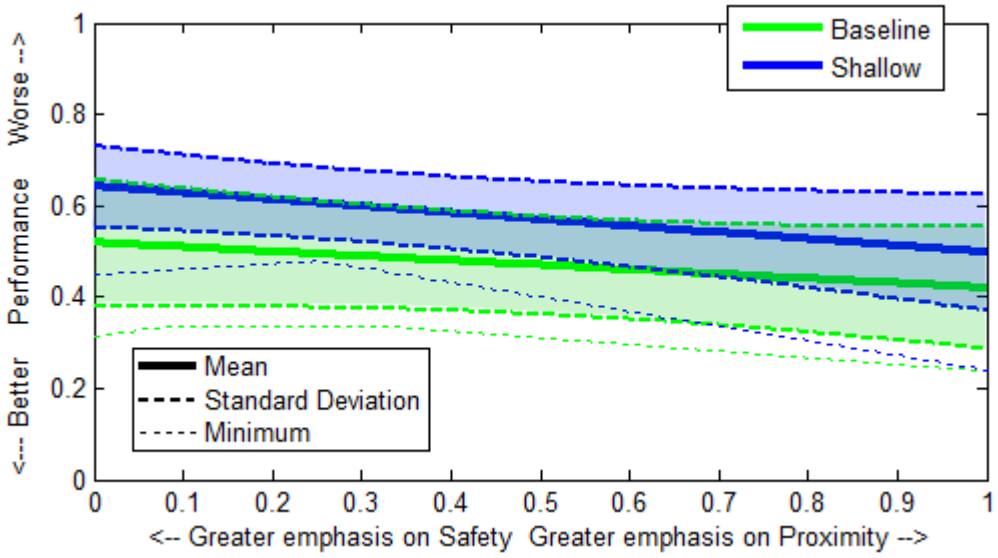


Figure 28: **Performance changes across trajectory profiles with variable weighting.**

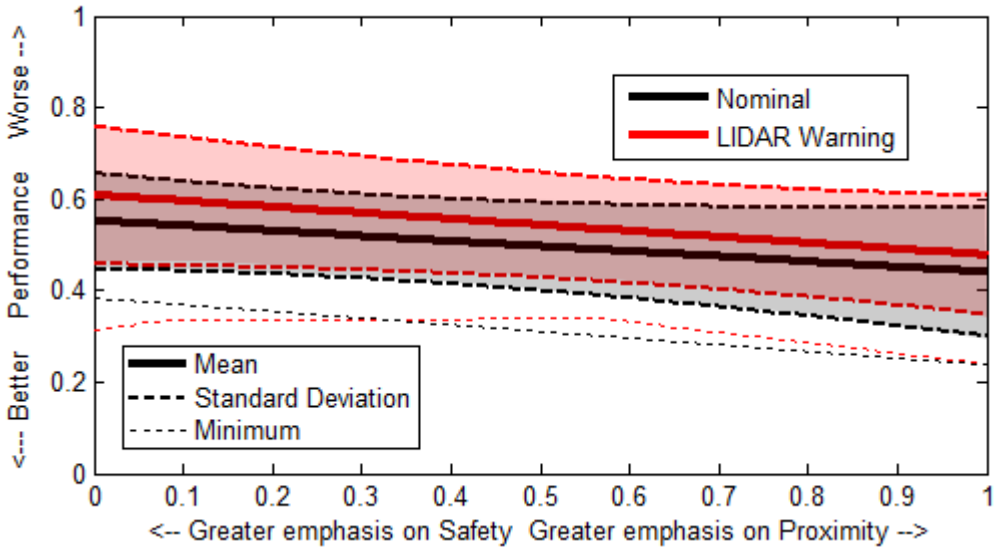


Figure 29: **Performance changes between different scenarios with variable weighting.**

each scenario. As evidenced in this figure, the changes in weighting result in varied performance. Participants appear to pick less safe sites during the LIDAR warning scenario, but their sites tend to be near the POI. Participants under the nominal scenario pick sites that are both equally acceptable in terms of safety and proximity to POI. Likewise, the LSS in both scenarios is effectively equivalent based on the site safety equality principle.

Figure 30 provides a comparison of participant performance (moderate and Apollo-like function allocations) and robotic function allocation performance against all scenarios. Each graph illustrates data of the participant within that specific scenario. Seven data points are included in the Apollo-like function allocation, and six are in the moderate function allocations. The robotic datum is one point, the top point suggested by the moderate automation sequence. A number of interesting trends are distinctive in this figure. There are several scenarios where the robotic function allocation clearly outperforms the participants. In the nominal, baseline cases, the robotic automation system routinely picks a site that is both safer and closer to the POI than the participants. This result implies that despite having the top site available to them, participants employing the moderate function allocation did not select it. However, in shallow trajectories, the participants select sites closer to the POI than the robotic automation. This result suggests that in short timelines,

participants are more likely to select an acceptable safe site that is close to the intended target. There is one instance where the participants outperform the robotic automation. In the baseline, LIDAR warning scenario, participants using the Apollo-like automation were capable of selecting a site better than the robotic automation. It may be that participants had sufficient time and control to determine a solution within the off-nominal scenario. The trends seen in the other three LIDAR warning plots illustrate that if the LIDAR sensor data is accurate, the automated landing site selection performs better than a human pilot. However, if the LIDAR data is inaccurate, the human pilot selects sites that are safer and closer to POI than the robotic case. The results presented in this section are sensitive to the quantitative scoring of landing site safety. As such, if weightings on specific factors such as terrain roughness or proximity to hazard change, or even using a probability density function of failure in lieu of a weighted sum, one should expect different results. Since the landing site score and the site option generation algorithm used in the experiment are not the same, it is possible for the automation systems in the robotic and moderate function allocations to select sites that are not the type sites within the landing area.

4.2.5 Impact of Search Method and Mental Modeling on Performance

Section 4.1.1 described the equal split of participants who used the areal and local search methods, and the near equal split of participants who used the reranking and eliminating methods. Table 17 compares the performance between the two groups, to determine if the search methods resulted in differential performance. Mann-Whitney tests illustrated that only time to complete was significantly impacted by search method, separated by function allocation. All other measures - LSS, Proximity to POI, Fuel, Workload - the strategies performed equally as well. Participants who used the areal search method, on average, took over 10 seconds longer than those who used the local method ($Z = -2.364, p < 0.018$). Participants who employed the reranking search method needed about 12 seconds more than those who used eliminating ($Z = -2.800, p < 0.005$). Boxplots of these distributions are illustrated in Figure 31.

A number of participants reported using the training photograph to develop a mental

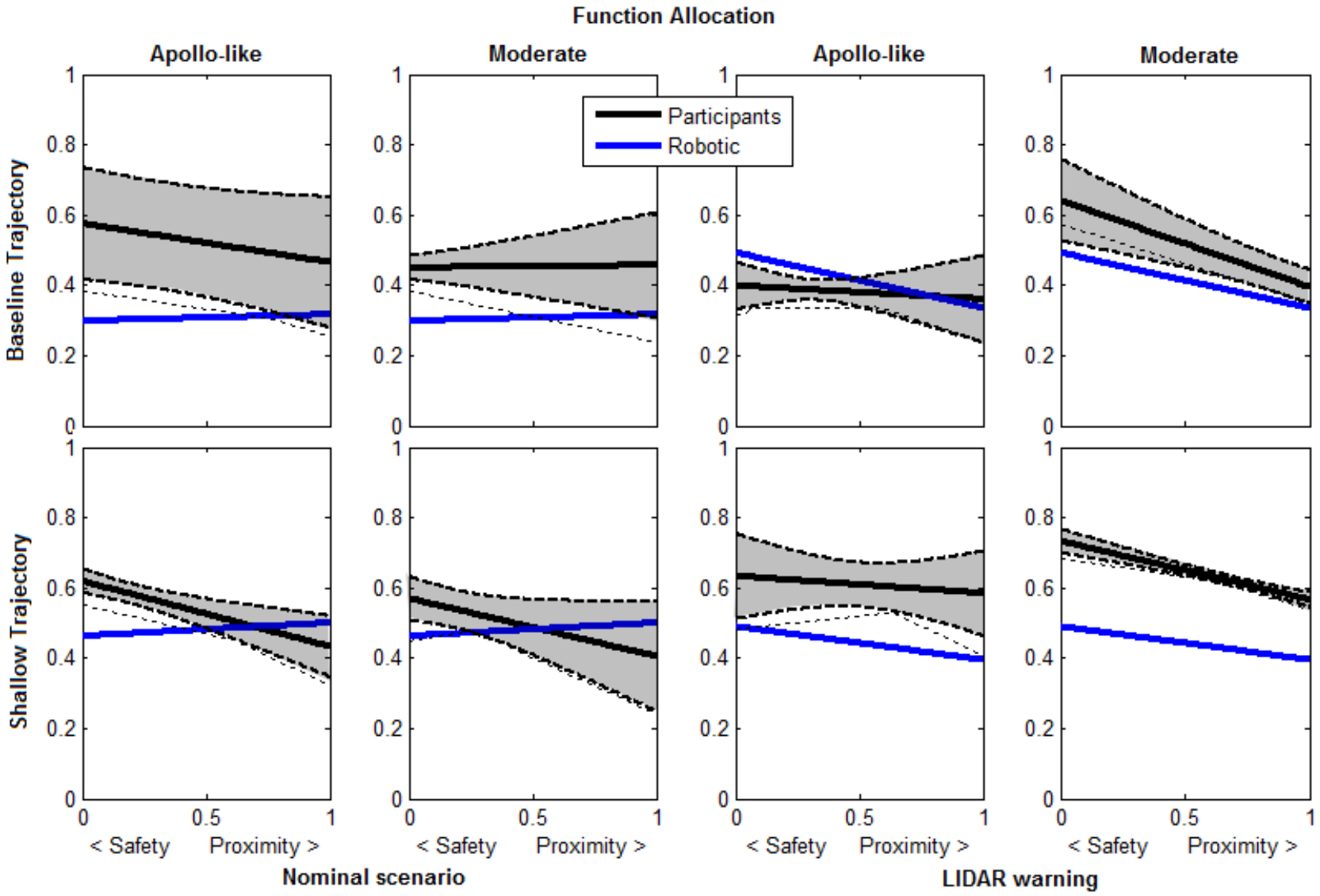


Figure 30: **Performance changes between different scenarios, trajectories, and function allocations.**

Table 17: **Comparison of areal/local and eliminating/reranking strategies.**

Apollo-like function allocation				
	Mean (std)			
Measure	**Areal**	**Local**	Z	p**-value**
T_C (s)	37.42 (12.64)	26.63 (17.46)	-2.339	0.017*
LSS	0.567 (0.153)	0.522 (0.120)	-0.664	0.514
Proximity to POI	0.456 (0.171)	0.451 (0.163)	-0.029	0.977
Fuel	531.73 (150.39)	537.25 (143.05)	-0.058	0.977
Workload	4.5 (2.355)	4.25 (0.452)	-0.181	0.887

Moderate function allocation				
	Mean (std)			
Measure	**Eliminating**	**Reranking**	Z	p**-value**
T_C (s)	20.85 (13.57)	29.48 (13.78)	-2.043	0.042*
LSS	0.602 (0.105)	0.618 (0.135)	-0.304	0.767
Proximity to POI	0.446 (0.119)	0.476 (0.100)	-0.630	0.537
Fuel	548.55 (135.65)	557.35 (135.56)	-0.743	0.478
Workload	3.50 (1.243)	2.88 (1.258)	-1.014	0.347

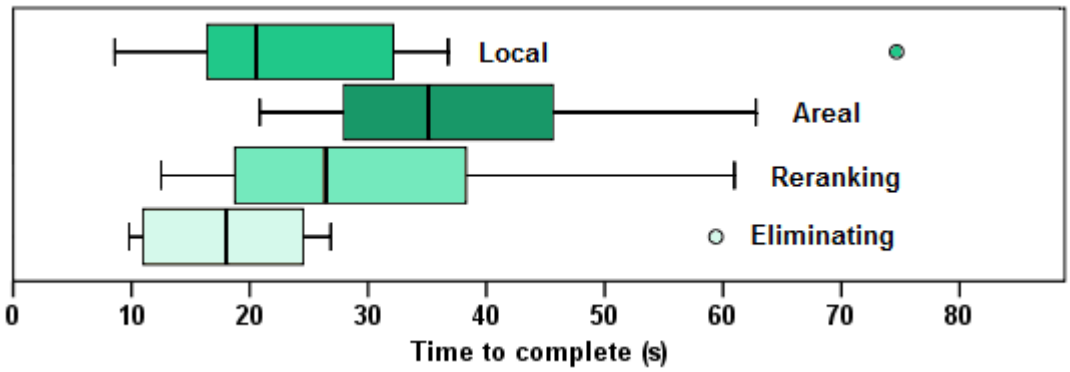


Figure 31: **Distribution of time to complete across different strategies.**

map of the landing area. A bivariate correlation shows that no correlation exists between participants who performed mental modeling and the reranking or eliminating search methods. However, a significant correlation ($r = 0.707, p < 0.00$) exists between areal and local search methods and mental modeling. Almost all participants who used the local search method developed expectations of the landing area and vice versa. The exception to this pattern is participant 107, a local searcher who formed a mental map for one of the six experimental runs. Participants using the Areal search method required 10s more than the local searchers. It may be that the first step of the Areal search method - finding a good region - is the analogous real-time action of mental modeling.

An ANOVA illustrates that the act of developing a mental model significantly effected the performance of both participants in the Apollo-like and moderate function allocations. The workload of the Apollo-like participants significantly increased (($F(1, 33) = 6.699, p < 0.014, \mu_{\text{workload-mmYes}} = 4.88$) for those who developed mental models compared to those who did not ($\mu_{\text{workload-mmNo}} = 3.75$). No other performance measure was effected. The additional workload may come from the act of searching for a specific landing site while integrating new information (LIDAR scan results) with preconceived perceptions of the terrain (mental map). For the participants using the moderate function allocation, both workload and time to complete were effected by mental modeling use. Workload significantly decreased (($F(1, 38) = 6.176, p < 0.017, \mu_{\text{workload-mmYes}} = 2.67$)) for those participants who developed mental models compared to those who did not ($\mu_{\text{workload-mmNo}} = 3.63$). Mental modeling may have allowed participants to quickly prefer or disregard sites in predefined good or bad areas. Timing was also significantly effected by mental modeling use. Participants who used mental modeling with the moderate function allocation performed the task significantly faster (($F(1, 38) = 20.402, p < 0.046, \mu_{\text{workload-mmYes}} = 29.796$)).

4.3 Discussion and Concluding Remarks

An experiment was designed to understand astronaut decision making during lunar landing and the associated changes in performance due to variations in function allocation, landing

trajectory profile, and sensor failure. Qualitative and statistical analyses of this experiment have elicited several major conclusions. First, depending on the function allocation, astronauts perform the landing point designation task in different but similar ways. Second, astronauts consistently used specific criteria with relative importance to generate their decisions. Third, the landing point designation task, with this hardware and software design, requires low workload regardless of the function allocation or mission characteristics, but gains in performance are achieved when astronauts are assigned responsibilities within the Apollo-like function allocation. Lastly, the astronaut experience, or flight background, significantly impacts performance.

Users of both the Apollo-like and moderate function allocations illustrated an equal split between holistic or individualistic search methods. Both the areal and reranking search methods evaluate all options on a whole (the map, the three sites) compared to the local and eliminating methods, which evaluated local sites (specific location, preferring the best or eliminating the worst). The task completion time is correlated with the search method used - the holistic methods required at least ten seconds more than the individualistic methods. However, neither search method provided an advantage over its counterpart within the same function allocation. Participants using the Apollo-like and moderate function allocation were also observed to use mental models, or pre-execution preparation, setting expectations for LIDAR scan results and candidate landing sites. While other mental models were likely used, this model was the only one measured. The frequency of mental modeling use was equal across the two search methods within the humans of the moderate function allocation. Interestingly, for those who were tasked with the Apollo-like automation, only astronauts who used the local search method also employed mental models of landing site expectations. It is possible that forming a mental model allows astronauts to skip the first step of areal searching, or the selection of good regions. Mental models allow for pre-identification of these sites and faster selection of the final landing site. However, forming mental models did not provide any gains in site selection or safety. The workload for Apollo-like local searchers was reported as greater than the Apollo-like Areal searchers. A possibility exists that the act of placing a landing site requires greater workload. As one participant commented,

"there's almost no place here [between the results of the LIDAR scan] where the circle will fit". This observation is reinforced by the behavior reported in the moderate function allocation. The workload for mental model users within the moderate function allocation actually decreases. Completion time also decreases for these same users. The limitation of three landing site options simplifies the decision making task and provides a visual anchor for the astronauts.

In addition to these observations with the search method, astronauts also used a series of decision making criteria to decide where to land. Generally, most astronauts used the same decision criteria, with differences accounting for the function allocation and the scenario. Astronauts were particularly focused on safety criteria, features that could be gleaned from the photograph of the expected landing area and the LIDAR scan results. Generally, opting to land away from and not on top of craters. The users of the Apollo-like function allocation, especially those using the areal search method, were more focused on least hazard density in lieu of the more localized away from craters and not within the landing site. Unsurprisingly, moderate function allocation users did not use least hazard density. The identification and limitation to three landing options allowed the users to focus in greater detail the hazards inside and within the neighborhood of the landing sites. Astronauts, on a whole, were also concerned with landing in areas of little visual information. Despite being removed from the actual execution of landing the vehicle, sites that were unfavorable for manual piloting were eliminated. Astronauts preferred sites that were well lit (only one astronaut had no preference within this category) and most aimed to land at sites of uniform color. Uniform color on the photograph was interpreted as flat terrain. While normally a reliable indicator of terrain grade, shadowing of lunar surfaces is heavily reliant on the approach angle and the sun angle. The trajectories were representative of optimal viewing conditions at the South Pole. It is quite possible to have landing sites that are uniformly colored and well lit, but on an incline. In the LIDAR off-nominal scenario, these visual cues were of greater importance than in the nominal scenario, an expected result due to the unreliability of the sensor data. Many participants, however, still opted to exclude LIDAR-identified hazardous areas, despite clear terrain directly beneath the overlay. These four criteria were generally

marked most important. Other criteria such as meeting expectations, fuel consumption, and proximity to POI were of less importance. Often, they were used as secondary criteria meant to distinguish between similar sites.

Overall, the landing point designation task was considered a low workload task. All astronauts had assumed at least one other astronaut would onboard the same crew, assisting him or her with the task of choosing a landing site. Up to three additional crewmembers were included in astronauts' assumptions of the cockpit environment. The roles were very clear, as universally described by the participants. The commander would be the sole decision maker and executor of the task. However, the other astronauts acted as sensory extensions, or decision support systems. These other crewmembers would call out state information during the allotted task time, recall procedures in off-nominal scenarios, watch for overlooked hazards, provide landing site suggestions, and maintain perspective of the mission objectives. In particular, these crewmembers would ensure that the commander would have only this task to focus on, regardless of the scenario. This comment was made by several astronauts and further confirmed with the minimal changes in workload between the nominal and LIDAR warning scenarios. The reported change in workload between scenarios, however, may be the result of the off-nominal scenario complexity. While the workload stayed fairly consistent between all the runs and the participants, aspects of performance were significantly effected by the function allocation, trajectory profile, and scenario. In general, astronauts chose safer sites with the Apollo-like automation compared to the moderate automation users. They completed the task proportionally to the time allotted. The additional thirty seconds due to the baseline trajectory resulted in safer sites chosen. Unsurprisingly, the safeness of the landing site was affected by the LIDAR warning scenarios. Astronauts felt more rushed and opted to land sooner rather than finding an optimally safe site. Additionally, in comparison to fully robotic systems (represented in the analysis as 0s designation time and selection of the highest ranked suggested landing site), neither the Apollo-like or moderate mode users performed any better. Even across variable weighting combinations of proximity to POI and landing site safety, the performance between robotic, moderate, and Apollo-like human-machine systems is equitable.

Pilots and commanders were observed to make minute changes until the very last possible designation moment, allowing for any possible incoming data. With the moderate automation, this behavior manifested as pressing the designation buttons at the final minute. In the Apollo-like function allocation, the selected site was continuously refined and adjusted until the end of the designation period. One participant, a commander of multiple Shuttle missions, declared the task impossible in the current configuration. Since the participant could not see the lander feet, the actions of the participant was effectively guesswork. He believed the role of the astronaut was to fine-tune and adjust the lander vehicle until the last minute, only building upon the majority of the work done by the automated landing system. Mission specialists, cockpit designers, and helicopter pilots were not observed to exhibit this behavior. Instead, they chose the landing site and allowed the vehicle to land at the intended site. Although no other participant explicitly asked for an allocation of responsibilities as this particular commander, many participants conveyed similar desires for continued involvement with the landing sequence.

CHAPTER V

COGNITIVE PROCESS MODEL DEVELOPMENT

The development of a cognitive process model allows for integration of such a subsystem with other working systems (e.g, automation, vehicle, landing trajectory, environment), to accurately represent a crewed landing system. This chapter discusses the development of a comprehensive landing model that includes all of these elements, but for the analysis conducted in Chapter 6, only the cognitive process model was used. The environment modeling has been previously explained in Chapter 3. The human cognitive process and the appropriate changes due to different function allocations is described in this chapter. This comprehensive landing model has been developed independently. The cognitive process model - accounting for both the moderate and Apollo-like function allocations - was developed as thesis work. The other models were built for NASA'S Autonomous Landing and Hazard Avoidance Technology (ALHAT) team and adapted for use in this thesis. Chapter 2 describes in greater detail the purposes of ALHAT.

For computational and development ease, the vehicle, vehicle dynamics, automation, and lunar environment models (to be referred collectively as the "ALHAT Simulation") operate independently from the cognitive process model. Thus a single data collection run contains three parts: 1) the ALHAT Simulation, then 2) the Cognitive Process Model, and 3) the ALHAT Simulation (Figure 32). The results of the cognitive process model are fed back to the ALHAT Simulation, wherein the ALHAT Simulation is modified and recompiled to account for these specific values. The cognitive process model reflects astronaut decision making when interacting with the automation system during an Apollo-like or moderate function allocation. It was not necessary to develop the decision making process for a fully automated system, as no humans are involved in the landing process during a robotic function allocation. Figure 32 illustrates this flow of information to each model. l.

The cognitive process model was developed from observations and analysis of the results

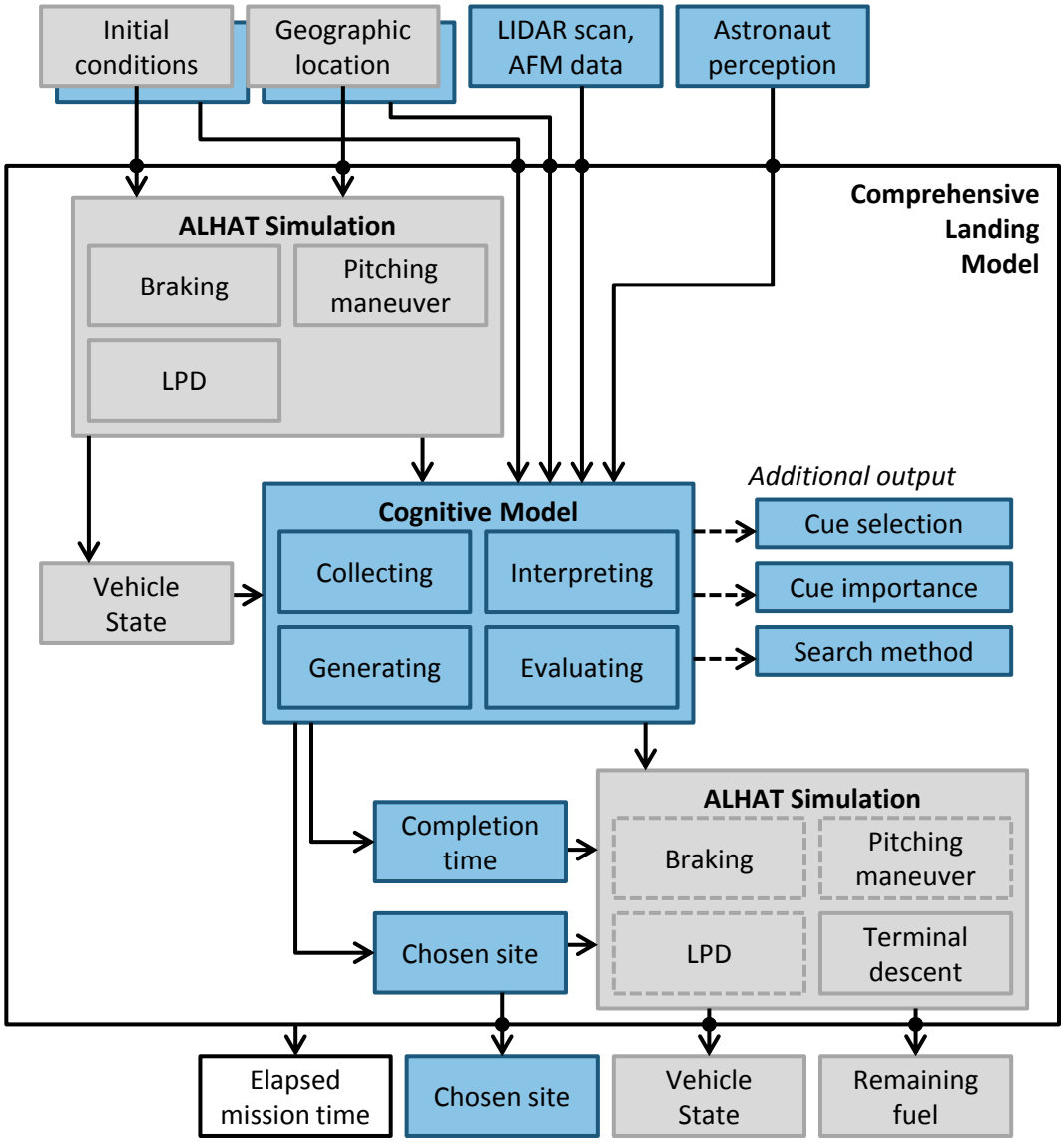


Figure 32: **Integration of ALHAT Simulation and cognitive process models.** The dashed lines represent aspects of the model that are executed, but not critical to the main operation. The boxes are colored according to their affiliation.

from the experiment, as described in Chapter 4. Twenty-eight data points were used to develop the interaction with the automation system of the moderate function allocation (i.e., moderate automation) and twenty-four for the Apollo-like function allocation (i.e., Apollo-like automation). The data points were evaluated to determine any outliers. Data points were considered outliers if they satisfied three conditions:

1. The datum point, or location of the chosen site, does not correspond with the self-reported criteria.

2. The chosen site is not near sites chosen by the other participants.

3. The chosen site is clearly not a good site, indicated by the availability of significantly better sites (greater than or equal to 0.1 of the score of the chosen site).

Using these criteria, only one site was declared an outlier and removed from consideration. Participant 110, while performing the LPD task under shallow, off-nominal conditions, selected a site on the edge of the LIDAR scan. Half of the site is within the 0.18 km $\times$ 0.18 km landing area. This site is permitted within the context of the scenario, but all other participants expressed a dislike of choosing sites close to the edge of the LIDAR scan, citing knowledge uncertainty. Participant 110, when asked about his decision making process during this run, stated that "I realized it wasn't the best decision but intuition... There was not enough time yes, during the sim[ulation], but I knew it was too late to make a change that big up to that point". There were clearly two other regions he could have landed at, one directly north of the southern crater, and one at the upper right corner of the map. No sites that were chosen by other participants are near this candidate site.

In both the Apollo-like and moderate automation scenarios, the participants were observed to behave in a manner similar to Klein's recognition primed decision model (RPDM) [31]. The premise of this model is that expert behavior relies on matching experiences (patterns) to real-time events. They quickly eliminate options or data that is irrelevant and only focus on the most critical aspects. Experts do not necessarily create and evaluate decision options holistically, rather, they select a suitable solution to the dynamic, time-pressured task. This solution may not necessarily be optimal, but it satisfies the criteria.

The participants demonstrated three distinct attributes of RPDM. First, ten of the thirteen of the participants, all astronauts, in the experiment were considered experts within the most related field. Within these astronauts, pilots and commanders were considered the closest equivalent to astronauts who would be trained to make decisions on landing location. These individuals performed differently, even from Mission Specialists and recent Astronaut Candidate graduates, by carefully diagnosing and monitoring the situation within the given experimental parameters. The landing site was adjusted accordingly. Second, half of the participants demonstrated pattern-matching, experience-relying behavior. Since none of the participants had previous experience landing on the Moon, photographs of the target landing area were presented prior to the experiment to simulate training and experience. The participants were not limited by time and could examine these photographs as long as they desired. About half of the participants stated that they developed mental maps prior to the each run and actively factored in this information into their decision making process. Third, half of the participants also eliminated or discard poor areas of the map, choosing to focus on regions they deemed preferable.

The participants that did not closely resemble RPDM followed more rational, analytical methods of decision making. These methods aim to optimize, rather than satisfy, by determining the best possible solution for a decision making event. Both RPDM and rational decision making features are represented in the models. However, as this model was designed as a cognitive process model, rather than a primitive model or cognitive architecture, algorithms for low-level functions such as perception and memory are not included. Rather, appropriate assumptions are made regarding these functions. All information regarding the task is assumed to be contained within working memory, with training elements drawn from long-term memory. Information from the display is pulled into short-term and working memory. The cognitive process model assumes the dominating source of perception is the visual channel, with the auditory and tactile channels not modeled. The cognitive process model is assumed to correctly sense, identify, and judge that the task is off-nominal.

The inputs to the cognitive process model were the landing site location (latitude and longitudinal coordinates), a flag indicating whether the scenario was nominal or off-nominal,

and the function allocation. The main outputs were the chosen landing site (latitudinal and longitudinal coordinates) and the time to complete, in seconds. Secondary outputs were internal factors that influenced the decision, such as the number and which decision making cues were used, the relative weighting of these cues, and the search method. Figure 33 presents an overview of the entire cognitive process model, including a summary of assumptions, theories, and empirical data usage.

5.1 Modeling the Reference Automation (Robotic)

Modeling a fully robotic automation system does not involve modeling human behavior. Subsequently, the model does not require human in the loop experimentation, unlike the other two cognitive process models proposed in this thesis. As described in Chapter 3, the landing point designation task can be broken down into five fundamental tasks: sensing, interpreting, creating, evaluating, and executing. In this function allocation, the automated system is handling all tasks.

The vehicle is based on a representative lunar lander as described by Davis et al. [65]. This lander, as seen in Figure 34 is derived from Apollo designs, bearing four legs and a central main engine used for both descent and ascent. The main engine is pump fed, oxygen-hydrogen propulsion and is assumed to be fixed and aligned with vehicle x axis. This engine has a maximum thrust of 357,081 N and a specific impulse of 440s. Sixteen reaction control system thrusters are located 2.6m below the top most part of the vehicle. The vehicle is approximately 10.5 m tall, a vehicle radius of 6.6m, four circular lander pads with radius 0.8m, for a total vehicle displacement approximated as a circle of 14.8m diameter.

The lander is equipped with an altimeter, velocimeter, and a LIDAR system to perform both terrain relative navigation and hazard detection and avoidance (HDA). For HDA, the LIDAR sensor is assumed to scan the terrain once, with the data feeding back to an autonomous flight manager for landing site selection.

Step 1: Sensing

In this step, the automated landing system uses the onboard sensors to determine the vehicle

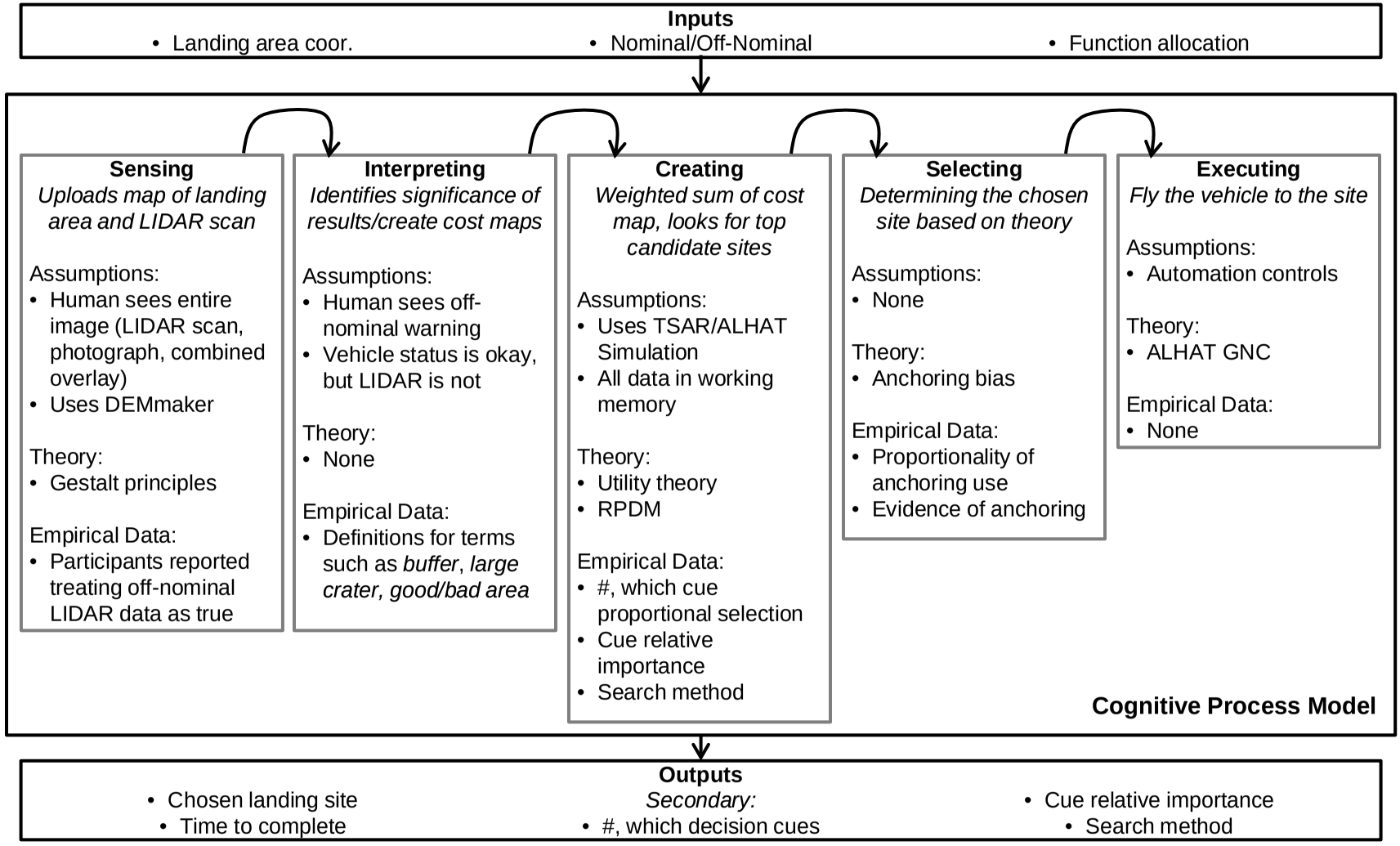


Figure 33: **Overview of the cognitive process model.** The cognitive process model takes three inputs and has two major outputs.

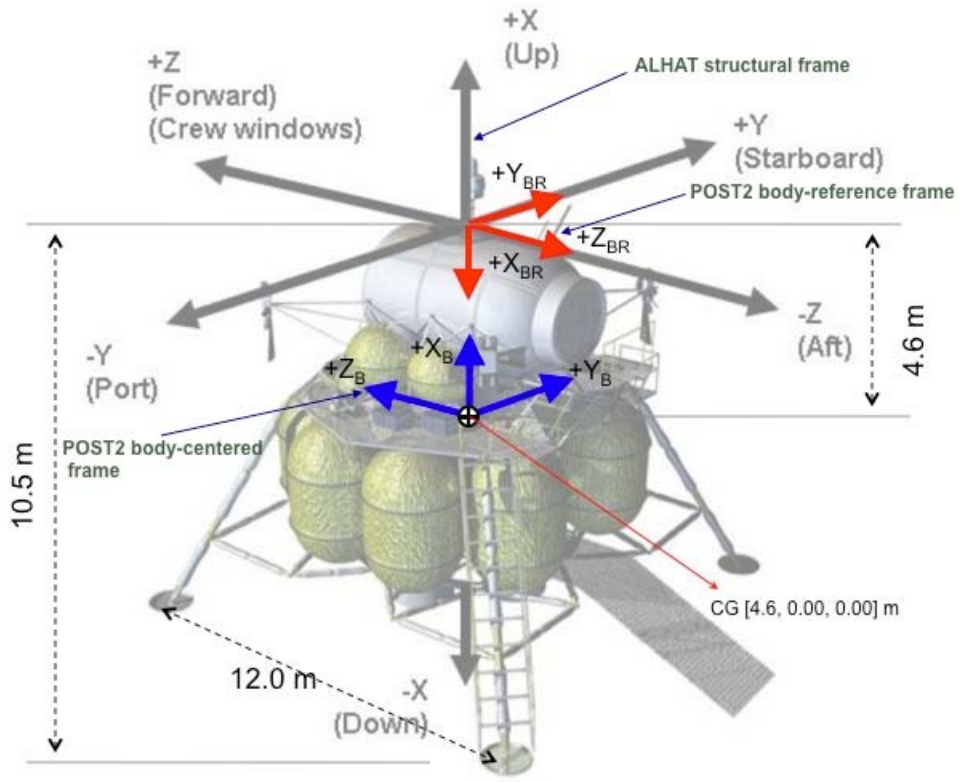


Figure 34: **Reference lunar landing vehicle.** Reprinted from Davis et al. [65]

state and the terrain characteristics of the expected landing area. At the end of the pitch up maneuver, the vehicle begins a 3-D Flash LIDAR scan of the terrain, in a 0.18 km $\times$ 0.18 km area, at a resolution of 0.2 m/pixels. There is also a star tracker, Doppler LIDAR velocimeter, an altimeter, and an inertial measurement unit (IMU) to provide information regarding the vehicle velocity, position, and orientation [95].

Step 2: Interpreting

The results of the LIDAR sensor scan are then processed to determine whether any terrain safety thresholds are violated. The terrain is marked hazardous if there are rocks and craters equal to or greater than 30 cm. Hazardous terrain also includes slopes greater than 10 degrees, as the lander will tip over [95].

Step 3: Creating

Following interpretation of the data, the automated landing system calculates cost maps based on proximity to the points of interest (centerpoint for this scenario), hazards, and change in velocity (i.e., fuel) costs based on the design of Cohanim and Collins [96], . Given a set of criteria weightings and a "spread" value (proximity of candidate sites to each other), a set of landing site options are generated from the weighted sum of all the cost maps. Candidate sites are ranked based on their cost map score, with the global optimum denoted with 1.

Step 4: Evaluating

In this step, the automated landing system always selects the top landing site.

Step 5: Execution

Lastly, the vehicle follows guidance laws based on the work of Fill [97]. The guidance algorithms used to control the vehicle during landing point designation attempt to fly the vehicle on a nominal trajectory to the start of the terminal descent phase, or strictly vertical descent. If the divert is completed within the area of the sensor scan, the vehicle does

not deviate substantially from the nominal path. There are two guidance laws - one for the lateral channel (crossrange and downrange) and one for the vertical channel (altitute descent rate). The lateral channel is governed by a bi-linear tangent steering law, in the form of a quadratic function to control of final position, velocity, and attitude (Equation 5, replicated from [97]. The coefficients of this function satisfy the boundary conditions.

$$\begin{aligned} a_c = \quad & P_n(t) + P_d(t) = c_0 + c_1 t + c_2 t^2 + \ldots c_n t^n \\ P_n(t) = \quad & C_{n,0} + C_{n,n} t^n \\ P_d(t) = \quad & C_{d,0} + c_{d,1} t + c_{d,2} t^2 \end{aligned} \tag{5}$$

P_n is the desired polynomial form on the nominal trajectory and P_d is the desired form on the divert. P_n is known prior to flight, but P_d must be solved for in real-time. The guidance law takes position and velocity targets (listed in Chapter 2) using Equation 6.

$$\begin{aligned} \Delta \vec{r_c} \quad &= \vec{r_T} - (\vec{r} + \vec{v} \cdot t_{go}) \\ \Delta \vec{v_c} \quad &= \vec{v_T} - \vec{v} \end{aligned} \tag{6}$$

where r, r_T are the current and targeted position; v and v_T are the current and targeted velocity; and t_{go} is the time to end of subphase. Fill describes the full derivation of the lateral channel in his paper [97]. The vertical channel control law is given by Equation 7 and is a proportional-derivative controller that maintains a desired descent rate.

$$a_x = a_d + G_p(v_x - v_d) \tag{7}$$

where v_d is the desired descent rate, a_d is the descent acceleration.

5.2 Modeling Decision Making within a Moderate Function Allocation

The moderate function allocation case in this thesis distributes the tasks of sensing and selecting to a joint human-automation team as seen in Chapter 3, Table 10. Fully automated tasks such as Executing are described in Section 5.1. A cognitive process model of human

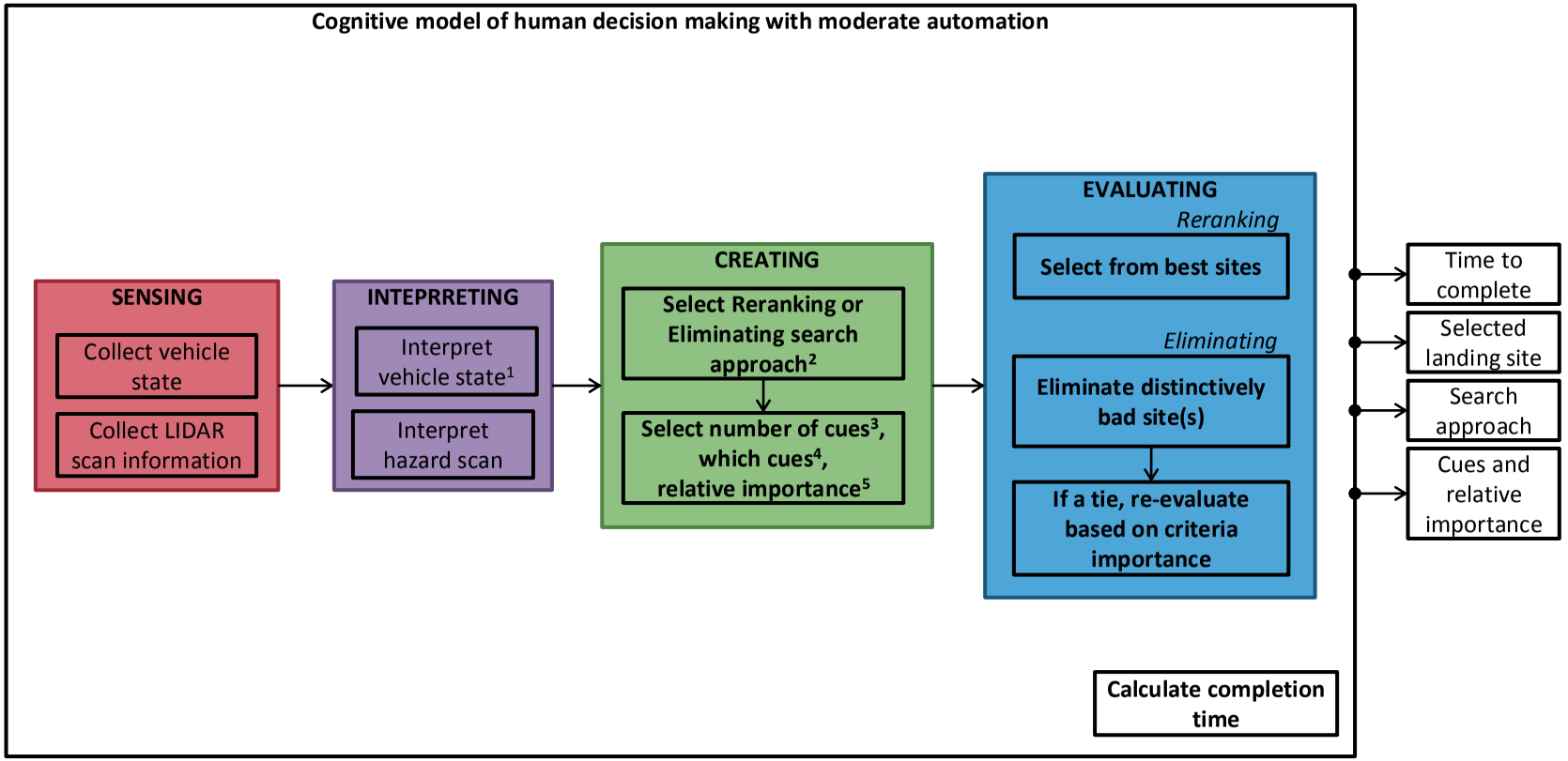
interaction with the automation was developed from the debriefing and experimental observations of participants. This section explains the algorithms and methods used to codify these actions and the validation process. Figure 35 illustrates the entire process.

Step 1: Sensing

Sensing consists of gathering information regarding the scenario and the environment. At this function allocation, the astronaut has several major sources of information: the photograph of the expected landing area (presented before and during each run), the views from the window, the results of the LIDAR scan, and the automation's suggested sites. As almost no participants actively factored in the window view into their decision making, this action was not included in the cognitive process model. The photograph of the expected landing area is included into the cognitive process model as an 8-bit grayscale image. The results of the LIDAR scan, especially in the off-nominal case, were not accepted literally by the participant. This information was used to develop a representation of the perceived hazards by astronauts. In some instances, astronauts were able to see past the LIDAR overlay and recognize that the LIDAR sensor had falsely identified a region to be hazardous. The perceived hazard visual is inputted to the cognitive process model as a matrix of binary values, with 1 representing a hazard and 0 as no hazard.

Step 2: Interpreting

Interpreting is the step following sensing and involves making a judgment, or an assessment, regarding the collected data. Generally, this judgment is simplified to two responses: acceptable and unacceptable or on a linear utility function. The collected data is often judged by the same criteria used for the actual decision making, or evaluating. In the case of LPD, the collected data is interpreted with respect to the candidate landing site (provided by automation or generated by the participant). There are ten attributes modeled, based on feedback from participants in the experiment: automation rank, away from craters, expectations of the expected landing site, projected fuel consumption, located near point of interest (POI), lowest hazard density, not in complete darkness, and uniform shading. Section 4



[1] Participants did not complete this task (because they were not piloting), this step is negligible.

[2] The number of participants who used the Reranking or Eliminating search approach is split almost even.

[3] The number of cues used by participants ranged from 1-5.

[4] Nine cues were used by the participants are inputs from the Training algorithm. This selection also includes frequent pairings observed in the experiment.

[5] This selection process also includes any notable patterns of importance assignments as observed in the experiment.

Figure 35: **Task diagram of human interaction with moderate automation.**

defines each of these attributes in greater detail. This section focuses on the codification of these cognitive assessments for the cognitive process model. For clarification, two terms are defined within the codification of the cognitive process model. Much of the information regarding the landing area was captured by matrices: row and column indexes corresponded to coordinates within the landing area (each unit representing 0.2m), and matrix values contained relevant information (e.g., distance of cell from another point; whether a hazard or not, etc). Therefore, a cell, i, refers to an individual entry in a matrix whereas a landing site refers to the area that is equivalent to a circle of a diameter 74 pixels.

Cue 1: Automation Rank

Participants using the moderate function allocation reported being influenced by the automation ranking of the landing sites. To score the three sites, the cognitive process model assigns inversely according to the suggested rank. That is, the top ranked site receives a score of 0, the middle ranked site receives a score 0.5, and the third ranked site receives a score of 1.

Cue 2: Away from Hazards

Participants reported searching for sites that were away from hazards, or large craters. When asked for additional clarification, participants defined a large crater as a crater that is approximately $0.25D$ of the landing site diameter. A suitable distance from a large crater, R_{buffer}, was defined as either one landing site diameter, or $1D$ away from the large crater, or none at all. Participants who reported using no buffer believed that the automated landing system was accurate and would not deviate beyond the landing site diameter. In the cognitive process model, both buffer distances were implemented, with a probability of $P(\text{no buffer}) = \text{`}0.5$ defining the frequency of using no buffer. To simulate the cognitive action of identifying a large crater, measuring the distance, and grading its acceptability, two additional graphics were developed to represent a human's sensing and interpretation of the terrain. The first graphic is an overlay that identifies large craters. This graphic is the same size as the map of the expected landing area and consists of two colors: a base color

to represent "no hazard" and a secondary color to mark "large crater/hazard". Any crater larger than or equal to $0.25D$ (19 pixels) is highlighted in the secondary color. The second graphic consists of points marking the edges of these large craters. The distance away from hazards is determined by calculating the distance between each of these points identifying the edge of a hazard and the candidate site. The minimum distance is considered the site's closest proximity to a hazard. The scoring for away from hazards is listed in Equation 8:

$$S_{\text{away}} = \begin{cases} \frac{R_{\text{buffer}} - \min(R_{i,j \to \text{Hazard}})}{R_{\text{buffer}}} & \min(R_{i,j \to \text{Hazard}}) \leq R_{\text{buffer}} \\ 1 & \min(R_{i,j \to \text{Hazard}}) > R_{\text{buffer}} \end{cases} \tag{8}$$

Future work may consider using a edge detection algorithm such as Hough transforms [98] to eliminate the additional step of self-identifying the large hazards.

Cue 3: Expectations

Calculating a score for each maps on the criterion of expectations requires experimental data. Several participants used the photograph of the expected landing area to develop mental models of where "good" areas and "bad" areas prior to the results of the LIDAR scan. Participants who performed this exercises were asked to identify on a map, where these good and bad areas were. These maps, as seen in Figures 16a, 16b, 16c, 16d, were processed into the cognitive process model. An identically sized matrix contained one of three values to indicate these responses, corresponding to the composite expectations graphic. Good areas were marked with a score of 0; neutral areas were marked with 0.5; and bad areas were marked with 1.

Cue 4: Fuel consumption

Participants reported using projected fuel consumption as a decision making criterion, but more as an approximation rather than the specific numerical value. No participant indicated using the digital readout of remaining kilograms of fuel. Rather, they approximated the fuel capacity based on the elliptical fuel ring overlay presented during the experiment.

To simulate this behavior, two response surface equations (RSEs) were developed to approximate the fuel required to land at a specific location relative to the center of the map. These equations are a function of time to complete (for the cognitive process model, this value is kept constant at 0 for each run, since it was reported that participants evaluate fuel consumption by location only, not including the temporal factor) and the x- and y-deviation from the center of the map. The RSEs were generated using experimental data, and diversions of 100-150m at 90 degree intervals at 20 second time steps. Forty-three data points were used for the baseline trajectory and 39 points in the shallow trajectory. The discrepancy is due to the 30 second increase in time in the baseline trajectory. The maximum fuel consumption possible for the baseline and shallow trajectories was 702 kg and 454 kg, and the minimums at 659.90 kg and 375.69 kg, respectively. To approximate a score for fuel consumption, the fuel consumption for each candidate cell was first calculated using Equation 9 and then scored based on 10.

$$m_{\text{fuel}} = \begin{cases} 659.90 + 1.16(t = 0s) + 0.03x + 0.00y & \text{if trajectory is baseline} \\ 375.69 + 4.21(t = 0s) + 0.05x + 0.01y & \text{if trajectory is shallow} \end{cases} \tag{9}$$

$$S_{\text{fuel}} = \begin{cases} |m_{\text{fuel}} - 659.90| \,/702 & \text{if trajectory is baseline} \\ |m_{\text{fuel}} - 375.69| \,/454 & \text{if trajectory is shallow} \end{cases} \tag{10}$$

Cue 5: Located Near POI

The proximity to the POI - always the center of the map - is calculated by the distance from the center to the candidate cell. Cells located closer to the center score better than those farther away. The farthest cell possible from the point of interest is $450\sqrt{2}$ m, or in the corner of the map. No other information is required to calculate this measure, so long the POI is well defined. Equation 11 gives the score for this criterion.

$$S_{\text{POI}} = (450\sqrt{2} - R_{i,j\rightarrow\text{POI}})/450\sqrt{2} \tag{11}$$

Cue 6: Lowest Hazard Density

This criterion is difficult to measure as it involves determining the population of hazardous regions around an ill-defined area around a specific landing site or potential candidate sites. For the cognitive process model, a basic approximation was used. Lowest hazard density requires identification of the hazards, as presented to the participants. From this data, the hazard density of each candidate cell is calculated by the number of surrounding cells that are identified as hazards.

$$\begin{aligned} \mathbb{1}_H \cong \quad & \text{1 if hazard, 0 if not hazard} \\ S_{\text{density}} = \quad & (\mathbb{1}_H(i-1,j-1) + \mathbb{1}_H(i-1,j) + \mathbb{1}_H(i-1,j+1) + \\ & \mathbb{1}_H(i,j-1) + \mathbb{1}_H(i,j+1) + \mathbb{1}_H(i+1,j-1) + \\ & \mathbb{1}_H(i+1,j) + \mathbb{1}_H(i+1,j+1))/8 \end{aligned} \tag{12}$$

Future work may consider determining the specific definition of hazard density considered by astronauts.

Cue 7: Not in Complete Darkness

This criterion is calculated with information only from the photograph of the expected landing area. This photograph, in black and white, converts to a 256 grayscale image, where low values correspond to the darker ranges. This range is representative of the intensity of a pixel. Candidate cells that contain 13.6% (36) intensity or less receive the worst scores. Intensities greater than this value receive a score based on Equation 13. This threshold was chosen based on empirical observations of participants' definition of darkness.

$$S_{\text{darkness}} = \begin{cases} 1 - \frac{C_{i,j}-36}{256-36} & \text{if } C_{i,j} > 36 \\ 1 & \text{if } C_{i,j} \leq 36 \end{cases} \tag{13}$$

Cue 8: No Hazards Within Landing Area

Calculating the percentage of hazards within a landing area required consideration of the actual area of each potential candidate site, rather than individual cells. Each landing site is

a circular region with a diameter corresponding to 74 pixels. As most calculations were done via matrices, an approximation of a square region 74 pixels wide was used. This landing site area was adjusted for candidate sites on the edges of the map. The hazard map presented to the participants was used to calculate the percentage of the landing site marked hazardous, as an approximation for hazards within landing area. The specific geometry of the hazards within the landing area was not considered. Equation 14 was used for scoring.

$$S_{\text{hazards in landing area}} = \frac{\sum_A (\mathbb{1}_H = 1)}{4^2} \tag{14}$$

Cue 9: Uniform Shading

This criterion is also dependent on the values ocntained within a landing area. A square region 74 pixels wide was used as an approximation for each cell calculation, with allowances made for cells located near the edges of the map. The photograph of the expected landing area is the only source of information for scoring the uniformity of the landing site. Uniform shading is approximated based on the number of pixels within the landing area that fall within ten percent of the average greyscale intensity. This threshold was approximated based on observations of participants during the experiment. Equation 15 describes the scoring relationship, where C_A is the intensity of the cells $C_{i,j}$ that are contained with in the landing area, A, and C_a is the subset of cells that are within 10% of the mean value.

$$\begin{aligned} \mu_{C_A} &= \frac{\sum_A C_{i,j}}{\|C_A\|} \\ C_a &\in [0.9\mu_{C_A}, 1.1\mu_{C_A}] \\ S_{\text{uniformity}} &= \|C_a\| \,/\, \|C_A\| \end{aligned} \tag{15}$$

Step 3: Creating

Creating is the generation of decision options that are later evaluated. As the moderate automation system creates three landing sites for the participant to choose from, the cognitive process model does not create decision options.

Step 4: Selecting

Participants using the moderate function allocation selected one of the three landing site options using a variable number of decision making cues and types of cues. They also performed one of two searching strategies, reranking or eliminating. Each method requires one round of evaluation. The cognitive process model randomly selects how many and which cues are used for each decision making process.

The number of cues used was randomly chosen for each moderate run using roulette wheel selection [99], with values proportioned based on Table 23 and Table 24. The relative importance of the cues is randomly selected, within the guidelines derived from selection patterns observed during the experiment (defined in Section 4.1.2). The relative importance is represented in the cognitive process model as a numerical weighting. Various combinations of weightings can be determined based on the number of cues selected. As the relationship between each cue can take only two values, a total number of combinations of weights per number of cues can be calculated. Table 18 illustrates the combination of weighting combinations per cue and an example of a numerical weighting scheme.

Table 18: **Relative importance combinations per number of cues used.**

Number of...		**Example of...**	
Cues	**Combinations**	**Relationship**	**Numerical Weighting**
1	$2^0 = 1$	C_1	1
2	$2^1 = 2$	$C_1 > C_2$	(1/3, 2/3)
3	$2^2 = 4$	$C_1 > C_2 = C_3$	(2/4, 1/4, 1/4)
4	$2^3 = 8$	$C_1 > C_2 > C_3 = C_4$	(3/7, 2/7, 1/7, 1/7)
5	$2^4 = 16$	$C_1 = C_2 > C_3 > C_4 = C_5$	(3/10, 3/10, 2/10, 1/10, 1/10)

A generalized equation for calculating numerical weightings is presented in Equation 16. One solves for x_L, the lowest weighting fraction assigned to the least important cue. The number of importance rankings can be defined as w (where 1 denotes the least important and w is the most important) and the number of cues N, and i the index of a specific cue. For example, the five cue relationship presented in Table 18 can be represented as $1 = 3x_L + 3x_L + 2x_L + x_L + x_L$, where w is equal to 3, N is equal to 5, there are two cues that are considered least important (C_4, C_5), two that are most important (C_1, C_2),

and one that is of average importance (C_3). The value for x is equal to 1/10.

$$\begin{aligned} N &= \sum i \\ 1 &= \sum_j^{C_i \in w=1} j(w=1)x_L + \sum_j^{C_i \in w=2} j(w=2)x_L + \cdots + \sum_j^{C_i \in w} jwx_L \end{aligned} \tag{16}$$

As described in Table 15, participants were observed to use some cues more often than others. Table 19 shows the distribution of number of cues used. The search method also influenced which cues were used. Table 20 shows the distribution of participants who used each cue. The user count reflects those who used that particular cue for that type of search method. For this model, the search method was randomly chosen based on an observed probability of 50% of astronauts using the Eliminating method, $P(\text{Eliminating}) = 0.5$.

Table 19: **Distribution of number of cues used with moderate function allocation.** These values are from experiment observations.

Scenario	Number of Cues				
	One	**Two**	**Three**	**Four**	**Five**
Baseline, Nominal	0 users	3	1	2	1
Shallow, Nominal	0	2	2	2	1
Baseline, LIDAR warning	0	1	1	4	1
Shallow, LIDAR warning	0	0	3	3	1

Table 20: **Distribution of which cues used with Moderate-like function allocation.** These values are from experiment observations.

	Nominal		**LIDAR**	
	Baseline	**Shallow**	**Baseline**	**Shallow**
Automation Rank	1 user	1	1	1
Away from Craters	5	5	7	7
Expectations	2	1	1	1
Fuel Consumption	0	0	0	0
Located near POI	2	2	2	2
Located Upper 2/3	0	0	0	0
Lowest Hazard Density	1	2	2	2
No Hazards within Landing Area	7	7	5	6
Not in Complete Darkness	1	1	3	2
Uniform Shading	4	4	6	6

The participants were observed, during the experiment, to clearly evaluate good and bad regions. However, participants did not differentiate, like other humans, below a certain

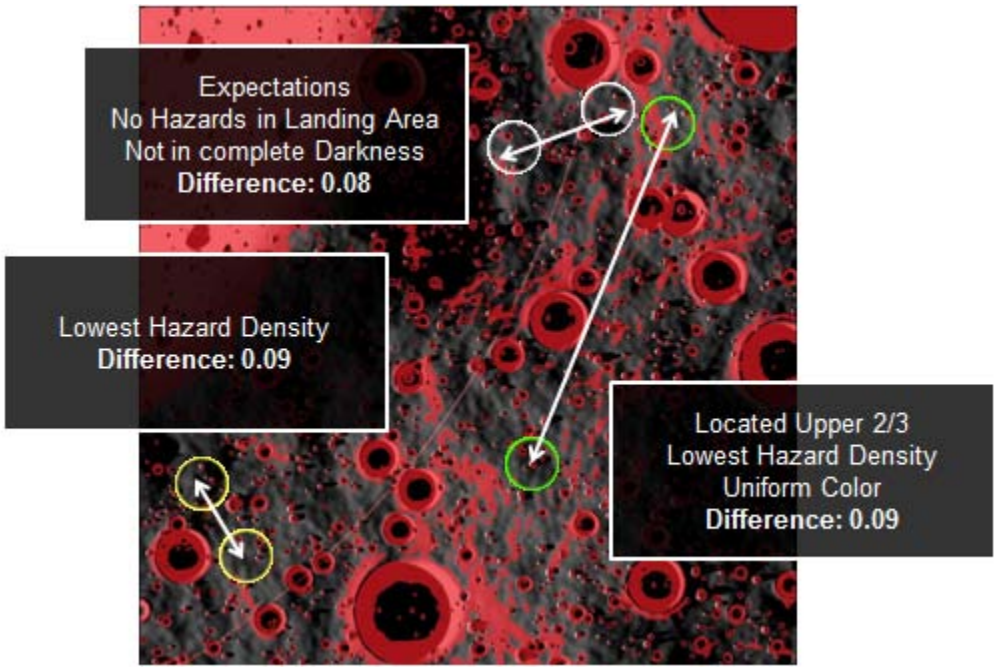


Figure 36: **Examples of sites within the equality threshold.** This equality threshold is based on observations of participants.

level of similarity. To model this rough equality between similar sites, the computational cognitive process model uses an equality threshold. The equality threshold is the numerical value of the cost score below astronauts did not differentiate between two sites. This value is about 0.1 and derived from observations of participants' chosen and considered candidate sites. Figure 36 illustrates some example sites that fall within this equality threshold.

The reranking search method is represented in the cognitive process model with a cost map. This cost map is a matrix of values from 0 (best) to 1 (worst) and is weighted score of each criterion and its respective weighting. The average score of the three landing sites are compared and checked for equality, i.e., having a cost score within 0.1. A site is randomly selected from those deemed to be equally good. However, the eliminating search method is less focused on score and more on eliminating an obviously poor site or selecting an obviously good site. The sites were compared based on the decision criteria used by the participant. The evaluation of each criterion follows the same schema, with slightly modifications accounting for the criterion type. A minimum, maximum, and median score is calculated between the three sites for a given criteria. The differences between the

minimum, maximum and median are calculated. If the median is closer to the minimum or maximum (defined as within the same quartile), then the incongruous site receives a score gain (0) or a score penalty (1). The gain or penalty is dependent on the objective function of that particular criterion. However, if the median resides in neither quartile, then no gains or penalties are imparted (each site receives a score of 0.5). This schema is used for each of the ten criteria.

Eliminating Cue 1: Automation Rank

As defined previously, points on the 0, 0.5, 1 scale were previously assigned to each of the sites. No further manipulation is necessary.

Eliminating Cue 2: Away from craters

The best site is the site with the largest buffer. The elimination schema is only used if the buffer range is greater than 10 pixels, or 2 meters. Otherwise, the buffers are indistinguishable.

Eliminating Cue 3: Expectations

The expectations map is input to the cognitive process model as an 8-bit grayscale image. The intensity within each landing site is averaged. If the average intensity is between 76 and 93 intensity (on a 256 scale), then the site is declared in a bad region. If the average intensity is between 109 and 154, then the site is declared in a good region. Other intensity values are considered located in neutral regions. The elimination schema is modified to account for the three possible values (good, bad, neutral) that each site can take. The full set of combinations is presented in Table 21

Eliminating Cue 4: Fuel consumption

The best site is the site with the minimum amount of fuel consumed.

Eliminating Cue 5: Located near POI

Table 21: **Scoring Heuristics for Astronauts' Mental Model of the Expected Landing Area.**

# of sites			
Good	**Bad**	**Neutral**	**Scoring Outcome**
1		2	good site receives score gain (0)
2	1		bad site receives score penalty (1)
	1	2	
2		1	neutral sites receive the penalty, the good site receives the gain.
1	1	1	good site receives gain, the bad site receives the penalty.
1	2		the bad sites receive the penalty, the good site receives the gain.
	2	1	the neutral site receives the gain, the bad sites receive the penalty.

The best site is the site closest to the center. The elimination schema is only used if the range is greater than 300 pixels, or 60 m, otherwise the distances are indistinguishable.

Eliminating Cue 6: Lowest hazard density

The best site is the site with the lowest hazard density. The elimination schema is only used if the range is greater than 0.01, otherwise the densities are indistinguishable.

Eliminating Cue 7: No hazards within landing area

The best site is the site with the least number of hazards within the landing area. The elimination schema is only used if the range is greater than 1%, otherwise the percentages are indistinguishable.

Eliminating Cue 8: Not in complete darkness

The best site is the site that is the lightest. The elimination schema is only used if the intensity difference is greater than 10. If only one site is considered dark, the other two sites receive score gains and the dark site receives a penalty.

Eliminating Cue 9: Uniform shading

The best site is the site with the most uniform shading. The elimination schema is only used

Table 22: **Distributions used to estimate time to complete for moderate function allocation users.**

Scenario	Distribution
Eliminating search method, baseline trajectory	$\mathcal{N}(22.79, 18.55)$
Eliminating search, shallow trajectory	$\mathcal{N}(18.91, 7.23)$
Reranking search, baseline trajectory	$\mathcal{N}(36.60, 14.43)$
Reranking search, shallow trajectory	$\mathcal{N}(22.36, 5.00)$

if the intensity difference is greater than 10, otherwise the intensities are indistinguishable.

After evaluating all of the sites, a weighted sum is calculated and the best scoring site is selected. If there is a tie, the cognitive process model uses the relative importance of the criterion as a tiebreaker. The site that individually scores higher for the most important criterion is selected. If a tie still exists after all criteria have been re-examined, a site is randomly chosen.

Step 5: Executing

The role of the automation within the moderate function allocation is to command the vehicle. The astronaut does not actually fly the vehicle to the intended site, rather, he gives the automation the location of the intended site. Subsequently, the fundamental task of executing is identical to the procedure explained in Section 5.1.

5.2.1 Timing

The cognitive process model was also designed to provide an approximation of the time to complete the LPD task. As described in Section 4.1.2, both the search method and the trajectory profile had a significant impact on the time to complete. The number of decision making cues or the specific types of decision making clues did not have a significant impact on time to complete. These guidelines were built into the cognitive process model. The time to complete the LPD task was approximated as a normal distribution. The relationships are described in Table 22.

It is worth noting that a few participants reported working faster during the LIDAR

sensor warning scenario, but they actually took the same amount of time. Since there was no correlation between the time to complete between both scenarios, this cue was not included in the cognitive process model. The participants likely felt an increase in time pressure due to the urgency of the scenario, but their behavior did not significantly change.

5.3 Modeling Astronaut Decision Making within an Apollo-like Function Allocation

Modeling human-automation function allocation when the human is engaged in several fundamental tasks is more difficult as the thought process is not easily observable. As a reminder in this thesis, the Apollo-like function allocation includes the human in all tasks except execution, or piloting the landing vehicle to the designated site.

Fully automated tasks such as Executing are described in Section 5.1. A cognitive process model of human interaction within the Apollo-like function allocation was developed from the debriefing and experimental observations of participants. This section explains the algorithms and methods used to codify these five fundamental actions and the validation process. Figure 37 illustrates the entire process.

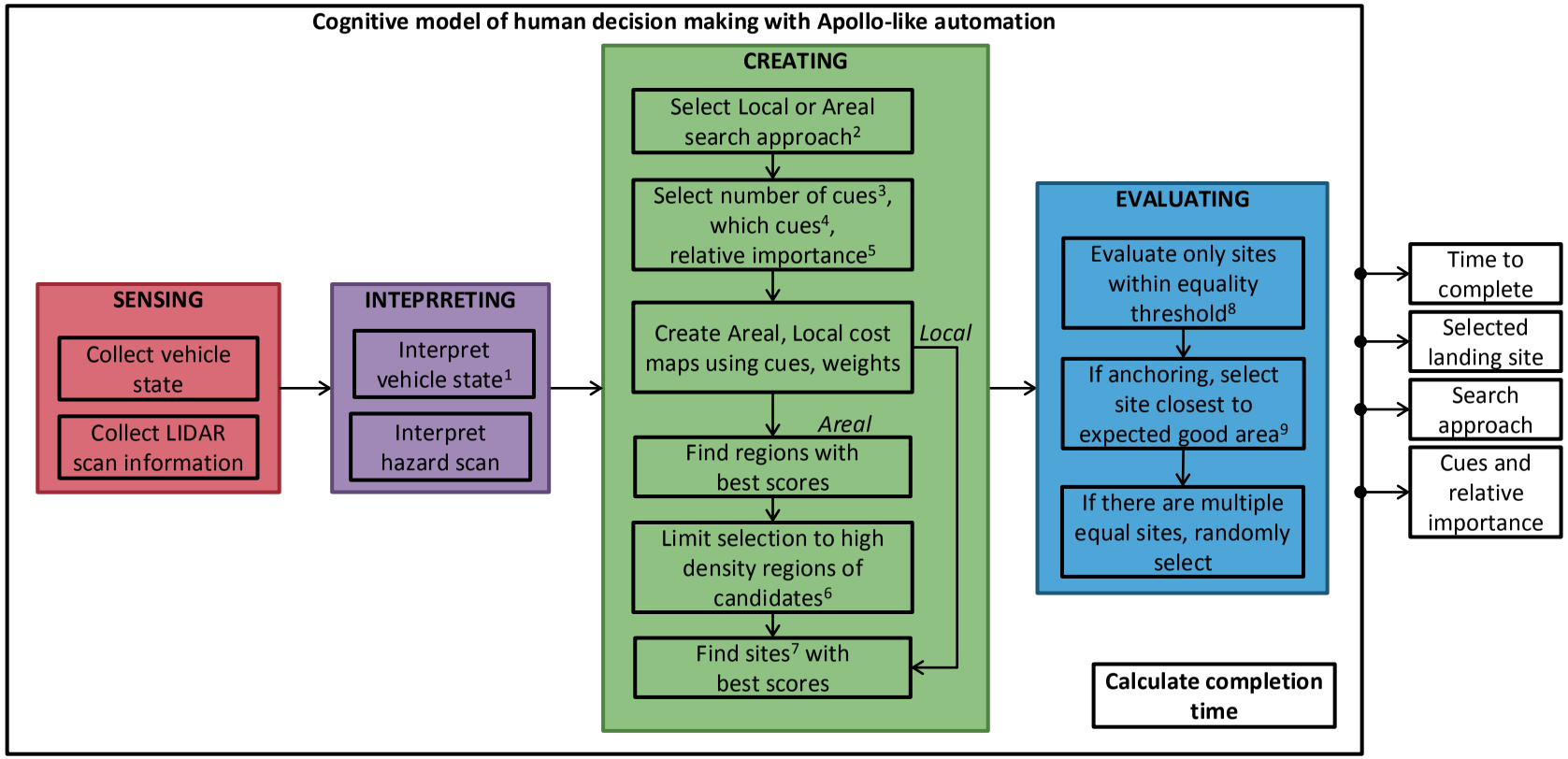
Step 1: Sensing

At this function allocation, the astronaut receives the same data as the moderate function allocation described in Section 5.2, except for the automation suggested sites. There is data on the vehicle position, velocity, orientation, and results from the LIDAR scan.

Step 2: Interpreting

Since the automation within the Apollo-like function allocation does not suggest landing sites, the computational cognitive process model must determine good landing sites as well as choose between them. Experiment results highlighted an additional cue that emerged from use of the Apollo-like automation: preference for sites that are located in the upper two-thirds of the map, due to an obscuration from the landing vehicle.

Cue 10: Located in the upper two-thirds of the map



[1] Participants did not complete this task (because they were not piloting), this step is negligible.

[2] The number of participants who used the Local or Areal search approach is split even.

[3] The number of cues used by participants ranged from 1-5.

[4] Nine cues were used by the participants are inputs from the Training algorithm. This selection also includes frequent pairings observed in the experiment.

[5] This selection process also includes any notable patterns of importance assignments as observed in the experiment.

[6] Participants reported narrowing sites down to good regions based on the Areal criteria, then selecting a final spot from the Local criteria.

[7] Landing area consists of a circle of *D* diameter, or 74 pixels.

[8] Candidate sites are not easily discernible if the score difference is less than 0.1

[9] Tversky and Kahneman's *anchoring* heuristic biased the selection process. The rate which this phenomenon occurs is derived from experimental results.

Figure 37: **Task diagram of human interaction with Apollo-like function allocation.**

Scoring for this cue is calculated by the y position. If the y position of the candidate cell is located in the upper two-third of the map, then that cell receives a 0 score. Otherwise, all cells in the bottom third of the map receive a score of 1. This scoring algorithm can be defined, with respect to distances relative to the POI, as in Equation 17. No other information is required to calculate this measure.

$$S_{\text{upper } 2/3} = \begin{cases} 0 & \text{if y} \leq \text{-1/6 m from center} \\ 1 & \text{if y ¡ -1/6 m from center} \end{cases} \tag{17}$$

Step 3: Creating

Participants using the Apollo-like automation used either the areal or local search methods to explore the map in order to create landing site options. The areal search method, as described in Section 4, is a two step searching mechanism that decomposes the landing area into smaller sites of interest. First, the astronaut focuses on good regions of the map - regions that score favorably on the list of criteria used in this round. Second, the astronaut selects a landing site from one of the good regions. The local search method is similar to the areal search method but is limited to one cost map, as only the low-level round is used. Astronauts used decision making cues in order to create one or more landing site options and the combinations of these cues differed based on the scenario type (nominal/off-nominal, baseline/shallow trajectory). The number of cues used to create and evaluate options also varied by participant and by landing scenario. To determine the number and which types of cues are used, the cognitive process model uses the same roulette wheel selection process as defined in Section 5.2. As described in Table 15 in Chapter 4.1.2, participants were observed to use some cues more often than others. Table 23 shows the distribution of number of cues used. It is clear that in each nominal scenario that most participants used three or more cues. This distribution is also used in determining the number of cues used in the low-level round of the areal search method. The cue selection process occurs in both the creating and evaluating subtasks, but are introduced in this section for ease of comprehension. The search method also influenced which cues were used. Table 24 shows the distribution of

participants who used each cue. The user count reflects those who used that particular cue for that type of search method. The cues used by local searchers and during the first, or regional round, of the areal search method are included under the 1st column; the second, or site round, cues used by the areal searchers are listed under the 2nd The search method was randomly chosen based on an observed probability of 50% of astronauts using the areal method, $P(\text{areal}) = 0.5$.

Table 23: **Distribution of number of cues used with Apollo-like function allocation.** These values are from experiment observations.

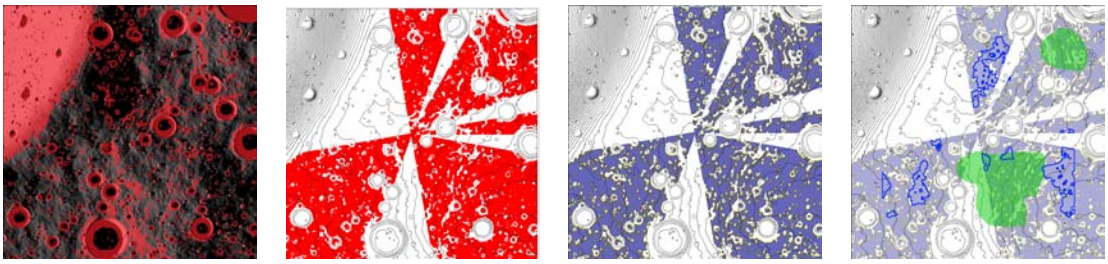
Scenario	Number of Cues				
	One	**Two**	**Three**	**Four**	**Five**
Baseline, Nominal	1	0	4	1	0
Shallow, Nominal	1	0	4	0	1
Baseline, LIDAR warning	1	1	2	1	1
Shallow, LIDAR warning	1	1	3	0	1

Table 24: **Distribution of which cues used with Apollo-like function allocation.** These values are from experiment observations.

	Nominal				**LIDAR**			
	Baseline		**Shallow**		**Baseline**		**Shallow**	
	1st	**2nd**	**1st**	**2nd**	**1st**	**2nd**	**1st**	**2nd**
Automation Rank	0 users	0	0	0	0	0	0	0
Away from Craters	0	1	1	2	1	2	1	1
Expectations	3	2	3	1	3	1	2	1
Fuel Consumption	0	0	1	1	1	1	1	0
Located near POI	2	1	2	1	1	0	1	0
Located Upper 2/3	1	0	1	0	1	0	1	0
Lowest Hazard Density	3	3	3	3	3	3	3	3
No Hazards within Landing Area	3	2	4	2	3	2	3	1
Not in Complete Darkness	1	1	1	0	2	0	3	0
Uniform Shading	2	2	2	1	3	2	3	3

5.3.1 Areal Search method

The cognitive process model mimics the cognitive process of a person using the areal search method by creating two cost maps, each representing the priorities of the astronaut during the high- and low-level rounds. Each cost map is a matrix of values from 0 (best) to 1 (worst) and is weighted score of each criterion and its respective weighting. The cognitive



(a) Map presented to participant.

(b) Results of areal, regional search.

(c) Results of areal, site search.

(d) Anchoring.

Figure 38: **Visualization of the areal search method.** These images visually represent intermediate steps within the cognitive process model. Fig. 38b is an example of the results of the areal, regional search. The red points represent the good regions based on the a set of decision making criteria. Fig. 38c illustrates the cost map based on the site search criteria, blue representing the best areas. This figure strongly resembles Fig. 38b, but several candidate sites have been eliminated as evidenced by the increased perforation in the same space. In cases where there are multiple good sites (outlined in blue), an anchoring bias can be factored in, as seen in Figure 38d. The backdrop is from Fig. 38c and the green represents the expected good areas identified by participants. If the anchoring bias is employed, the site closest to the expected good area is chosen as the cognitive process model's landing site.

process model begins searching for good regions by collecting all cells within an equality threshold [1] of the best score on the cost map. Areas with the highest densities of "good" cells are selected for further evaluation. This area selection process is approximated by dividing the cost map into thirty-six equal pie sections and claiming the sections containing the majority of these good cells. The selection of a landing site is limited to these areas culled from the first round, or high-level search. This process is illustrated in Figure 38.

In the second round, the cognitive process model uses the cost map generated from the cues and relative importance of the low-level search. The corresponding scores of the candidate cells from the previous round are collected from this second cost map. The scores of all the neighboring cells consisting of the landing area are averaged for a landing site score. An averaged score is used rather than the score of the candidate cell to accurately reflect the site scoring completed by participants. Sites not completely contained within the map or are more than 20% over a large crater are discarded. The participants reported avoiding the edges of the LIDAR scan as the information is incomplete. The crater overlap

[1] Definition provided in Section 5.2.

threshold was determined based on experimental data. The final sites are ranked in order of best score (lowest to highest). Candidate sites created during this process are defined as sites within the equality threshold of the top scoring site. These sites are saved for final evaluation.

5.3.2 Local Search Method

The cognitive process model collects all cells within an equality threshold of the best score on the cost map. Once the obvious violators have been discarded (too close to edge of LIDAR scan, overlapping with craters), this list of final candidates is ranked in order of best score. Candidate sites falling within the equality threshold of the top scoring site are saved for final evaluation.

Step 4: Selecting

The decision making process, regardless of the type or theory used, generally involves some form of evaluation. The evaluation process, as observed by the researchers, seemed to occur in one of two ways. The astronauts often chose one of the acceptable sites generated through the creating process (with the possibility of additional, minute refinements) or they tended to select acceptable sites that were near expected good regions, determined prior to the run. This second phenomenon is known as anchoring, defined by Tversky and Kahnemann as "[estimations made by people] by starting from an initial value that is adjusted to yield the final answer" [100]. Figure 38d illustrates this anchoring process. However, anchoring was not observed universally with all participants. The probability of anchoring, or P(anchoring) was determined to be 4/6. This number reflects the four participants who reported setting expectations prior to the run.

If anchoring is in use, the cognitive process model performs additional steps during evaluation. First, the closest distance between all of the candidate sites and the good areas of the expectations map is calculated. Second, the closest distance is selected as the final landing spot. If multiple sites are within the same proximity (e.g., many sites tend to be within the good area), then a final site is randomly selected. If anchoring is not in use, then

Table 25: **Distributions used to estimate time to complete for Apollo-like automation users.**

Scenario	Distribution
Areal search, baseline trajectory	$\mathcal{N}(47.13, 9.89)$
Areal search, shallow trajectory	$\mathcal{N}(27.70, 5.22)$
Local search, baseline trajectory	$\mathcal{N}(29.68, 24.33)$
Local search, shallow trajectory	$\mathcal{N}(23.58, 7.51)$

the cognitive process model randomly chooses between all of the candidate sites created in the previous step. The cognitive process model randomly chooses between these sites.

Step 5: Executing

The role of the automation system within the Apollo-like function allocation is to execute the vehicle. The astronaut does not actually fly the vehicle to the intended site, rather, he gives the automation the location of the intended site. Subsequently, the fundamental task of executing is identical to the procedure previously used in the moderate function allocation.

5.3.3 Timing

The cognitive process model was also designed to provide an approximation of the time to complete the LPD task. As described in Chapter 4, both the search method and the trajectory profile had a significant impact on the time to complete. The number of decision making cues or the specific types of decision making clues did not have a significant impact on time to complete. These guidelines were built into the cognitive process model. The time to complete the LPD task was approximated as a normal distribution. The relationships are described in Table 25.

It is worth noting that a few participants reported working faster during the LIDAR sensor warning scenario. Since there was no correlation between the time to complete between both scenarios, this cue was not included in the cognitive process model. The participants likely felt an increase in time pressure due to the urgency of the scenario, but their behavior did not significantly change.

5.4 Validation

Validating a computational model, especially one built from empirical observations, generally requires a subset of data to test the model accuracy. Or, if a subset is not possible, statistical methods such as bootstrapping are used to derive statistical parameters from a very small sample of data. Neither of these methods were feasible for this cognitive process model. The cognitive process model of astronaut decision making was derived from 28 data points for the aspect with the moderate function allocation, and 23 data points for the Apollo-like function allocation. Seven data points for each scenario (two trajectories, two scenario types) were available for the moderate function allocation, whereas only six points per scenario were available for the Apollo-like function allocation. With such a small dataset, it was not possible to reserve a portion of the data for validation, as all data points were necessary for model generation. Additionally, bootstrapping was not possible as many aspects of the cognitive process model do not follow known distributions.

This cognitive process model is primarily built from the debriefing sessions with the participants. The same participants were used to build and validate the model, but the specific information is different. These debriefing sessions provided qualitative information regarding the overall strategy, the cue usage and relative weightings. The model was validated on the quantitative results of the experiment, on the sites selected from the same individuals whose cue usage and preferences provided the input data.

Assuming the cognitive process model selects how many, which cues, which search method at the same proportionality observed during the experiment and that the time to complete value is a reasonable approximation based on the timing scores of the participants, the model should satisfy two other conditions:

1. **The cognitive process model should select the same site chosen by each participant during the experiment.** The candidate sites selected by the cognitive process model during the creating process should match the participant's chosen site, when using the same criteria, relative importance, and search method as reported.

2. **The cognitive process model, when sampled randomly, should select sites**

physically near those chosen by the other astronauts. The cognitive process model was run 1000 times and the spread of the selected sites was compared against the participants' chosen sites.

Condition 1 was examined by running the cognitive process model with the same cue combinations, relative importance, and search method as described by the participant. The cognitive process model could produce three quality levels of results, with respect to matching the participant's chosen site. These levels - correct detection, missed detection, or false alarm - are based on signal detection theory [101]. Correct rejection cannot be defined within this context as participants were never explicitly asked to reject sites, nor was that the intended purpose of the cognitive process model. A correct detection is when the cognitive process model's selected site is the same site as the participant's chosen site. That is, the top scoring site as selected by the cognitive process model is the same as the participant's chosen. A missed detection occurs when the cognitive process model's selected site is not the same as the participant's chosen site, but instead, the cognitive process model's second ranked site matches. A false alarm occurs when the cognitive process model's selected and second ranked sites do not match the participant's chosen site.

Using these definitions of result quality, each participant's individual runs were executed. After each run, the candidate sites were plotted and compared to the participant's chosen site. Of the 28 moderate function allocation runs, the cognitive process model produced 27 correct detections, 1 missed detection, and 0 false alarms. The missed detection occurred during the shallow LIDAR warning scenario.

This same procedure was used to validate the aspect involving the Apollo-like function allocation, with a few differences. First, twenty-three runs were used to validate Condition 1. These runs were produced using the same decision making criteria as the twenty-three experimental points used to generate the cognitive process model. Figures 56-59 present the visualization of all the candidate sites and the participant's chosen site for each run. Of these runs, the cognitive process model produced 20 correct detections, 3 missed detections, and 0 false alarms. All three missed detections occurred during shallow trajectories; two happened during the nominal trajectory.

Condition 2 was evaluated by running the model 1000 times. Each run consisted of a random combination of decision making cues, relative importance, and search method. The selected sites were plotted with the participants' corresponding selection. The distribution of selected sites matches, as seen in Figure 39. The distribution of how many, which cues, search methods, and time to complete approximations also match (Figures 50, 51, 52). The cognitive process model also selected sites that were not chosen by the astronauts, suggesting that under different cue usages and their relative importance, sites may be more or less viable.

As for the Apollo-like function allocation, the majority of the runs populated by the cognitive process model matched common regions selected by the astronauts (Figure 40). The cognitive process model also selected sites in regions not chosen by the astronauts, suggesting that under different cue usages and their relative importance, other regions may be more or less viable.

The cognitive process model provides limited predictive capability. Although the types of cues, relative weightings, and search method are randomly assigned, the cognitive process model can accept these values as inputs, in addition to the original three. This capability allows mission designers to predict how these decision cues contribute to the chosen landing site. Information analyses regarding availability or reliability of data are possible, to examine performance in off-nominal situations (as demonstrated with the LIDAR warning scenario). Likewise, mission designers can determine the spread of possible landing sites given a specific landing area.

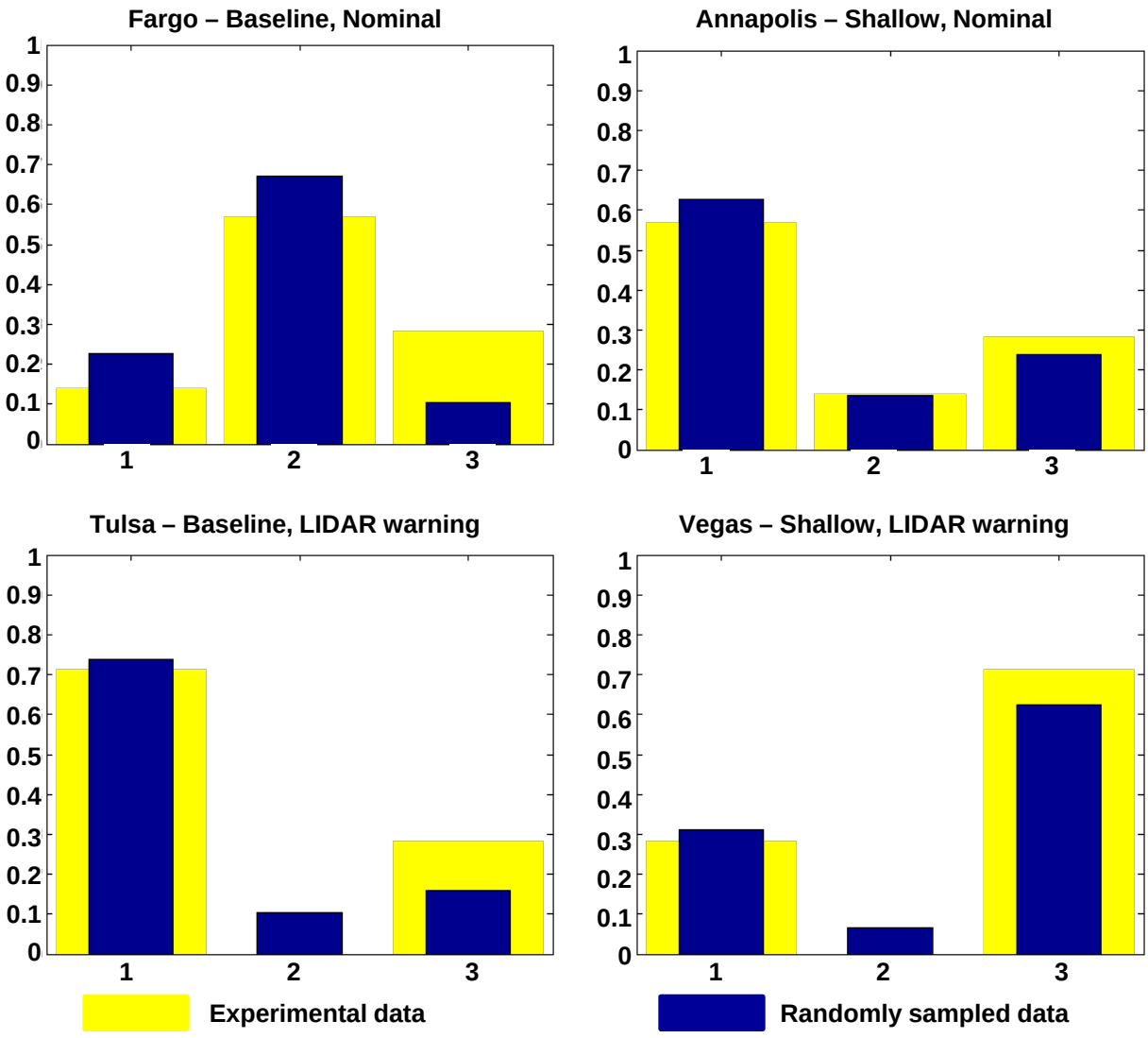


Figure 39: **Distribution 1,000 of randomly generated landing site decisions in compared with experiment results.** Fargo, Annapolis, Tulsa, and Vegas are nicknames used in this thesis for the landing sites on the south pole of the Moon.

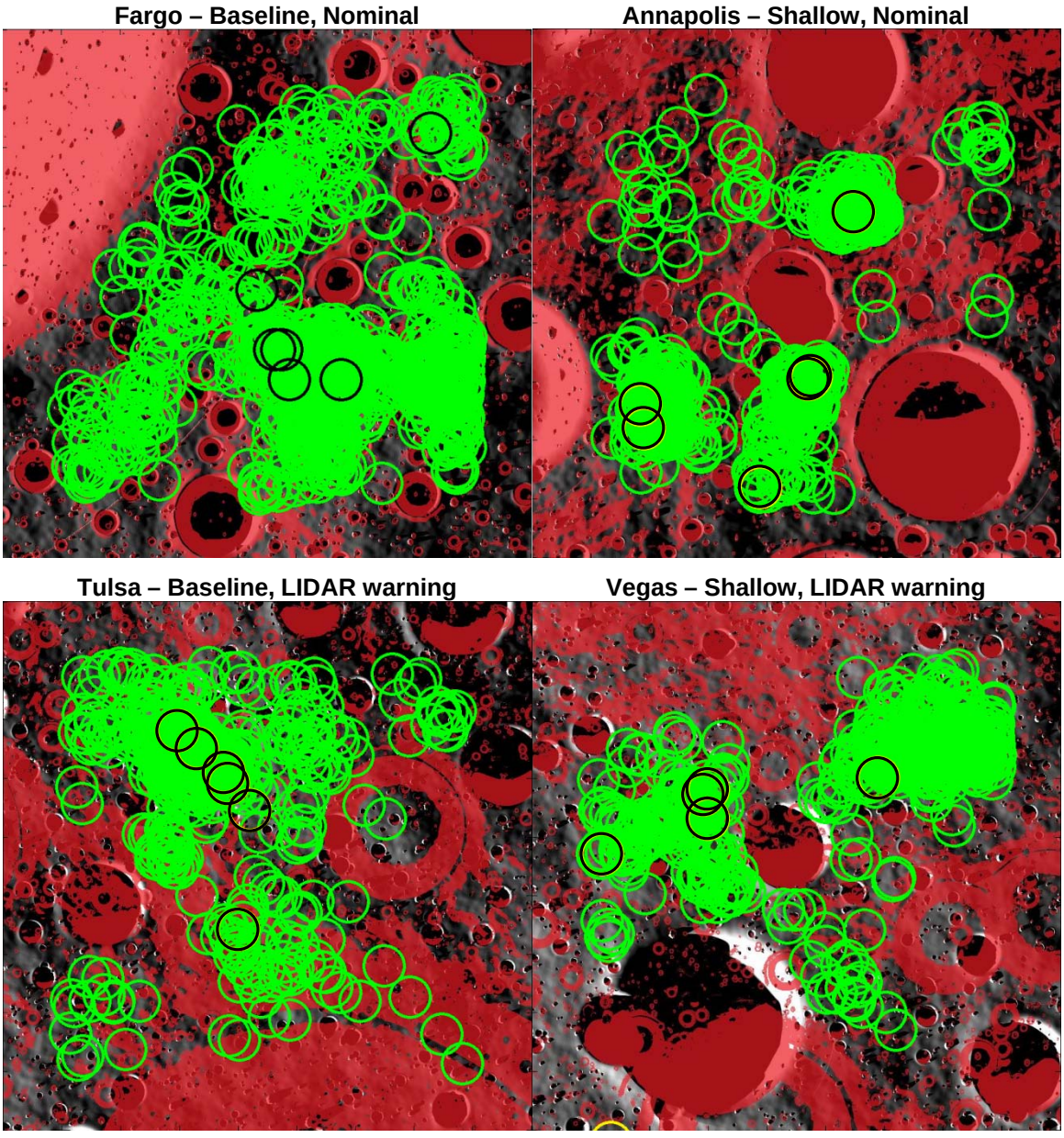


Figure 40: **Distribution of 1,000 randomly generated landing site decisions compared with experiment results.**

CHAPTER VI

SUGGESTED AUTOMATION RESPONSIBILITIES AND CREW TRAINING OBJECTIVES

Section 3 and Chapter 4 described the design, execution, and analysis of an experiment intended to examine the impact of function allocation, trajectory, and scenario on human-automation interaction. During this experiment, the participants were also asked to provide feedback on the cockpit layout, the display suite, and the type of information available to the participants. This chapter presents the results of the qualitative analysis and proposed requirements for cockpit design and training. Additionally, a normative model for automation interaction is presented for future crewed missions. This normative model is the basis of automation system requirements necessary to support future crewed landing point designation.

The suggestions and feedback provided by the participants touch on three areas of overall cockpit design: information needs, method of representation, cockpit layout. Each of these three areas are discussed, including associated system requirements within each category. Some of the requirements are derived directly from participant feedback; others are from the observed collective behavior and the results of the cognitive process model (Section 6.4). In addition to providing feedback in these areas, participants also discussed the desired function allocation during landing point designation (Table 26). The feedback from the participants provides perspective on the original task decomposition of Chapter 3, resulting in an updated task decomposition and function allocation scheme (Section 6.6). This analysis also draws on literature to support suggested design heuristics. As prescribed by Chapanis [11], these design requirements are written to effectively incorporate human factors issues into the design process in the standard language used by all other subsystems designers. These requirements represent the work of just the human factors specialist,

prior to consultation with other specialists such as cockpit display designers and ergonomicists. Such specialists would use handbooks and regulations such as MIL-STD-1472 [45], MIL-HDBK-761 [44], and NASA-STD-3001 [41], to mature and detail requirements to the necessary project. Lastly, observations of participant behavior and biases are compiled to suggest additional topics for future training programs for landing point designation.

6.1 Information needs

Analysis of the information, display, and cockpit suggestions provided by the participants reveals a number of design requirements that can be passed to mission design. Section 6.4 contains all of these suggestions. The flow of these requirements follow the structure of this chapter, with three categories: information needs, method of representation, and cockpit layout.

Overall, the composite sum of information, from the reference displays found in literature, presented to the participants during the experiment was sufficient to complete Landing Point Designation (LPD). However, participants requested additional information to support this task, improving confidence in the performance of the system and the decision of the individual. In the experiment, the simulation was stopped after the allotted decision-making period had expired, with confirmation that the vehicle successfully touched down at the exact site at each designation. One participant repeatedly commented on the need to see the vehicle landing pads, "to have a good feel for where things are. Just like a runway, I want to see the landing zone." He felt that the task was impossible, or even a "negative" task, as "I just make a decision and oops, I'm out of it. I almost give up as the operator, I have no other inputs, you're not giving me the ability to fix what I think is the most critical portion of the task, and that's physically landing on the planet." He believed that "the most important aspects [of LPD] are accurately predicting where the vehicle is going to be at", and the LIght Detection and Radar (LIDAR) sensor was insufficient, especially as it had a chance for failure. Another participant agreed with this statement by expressing his belief that the pilot's greatest contribution came from placing the lander pads between the rocks and craters, rather than selecting a generalized landing zone. Requirements 1.2.1, 1.3.1,

and the respective sub-requirements are derived from this observation. Other participants also touched upon this theme of LIDAR insufficiency. Several participants wanted multiple LIDAR scans, to continue getting real-time updates and to correct itself (Req. 1.1). Additionally, they requested LIDAR scans of the landing sites itself (the three designated by the automation, or one designated by the participant, Req. 1.1.1). While participants did not provide a specific LIDAR scan resolution for the sites themselves, the suggested resolution is based on the lander footprint size. In the experiment, participants defined large hazards as those within a quarter of the size of the landing footprint diameter. Since the main reference point during this exercise is the lander pads themselves, it is assumed that a resolution of one pixel should capture a distance equal to 0.25 of the lander pad diameter. If additional scans were not possible, then a camera tracking the landing site or a better look angle through the camera or window was a necessity (Req. 1.2.1). Participants were observed to be frustrated by the disappearance of the landing site as the vehicle drew closer to the ground, with one participant actively refusing to select a site in the lower third of the display, as to avoid losing out on such critical information. A user-controlled gimbaled camera was suggested, especially if hover capability was provided to the crew, but with the caveat that controlling the camera angle was another task for the operator.

The participants also commented on wanting control over the types of information presented. Almost half of the participants asked specifically for a declutter button, to remove the results of the LIDAR scan, especially if results were poor (Req. 1.1.2). One participant noted that having to sort through erroneous LIDAR results increased his overall workload: "I'm trying to declutter [the display information] myself - [this secondary task] became a distraction... If the LIDAR can sense the issue, give me the warning, don't give me the data... It cuts down on my workload". Several participants also asked for gradient hazard shading or two levels of hazard identification: "I could potentially get scared away from this area, because it's showing a lot of red but if you take a look at the landing zone, well it doesn't look all that scary, but it might be seeing stones and all" (Req. 1.1.3). Another suggestion involved reducing the pilot's workload by highlighting feasible landing spots in lieu of hazards (Req. 1.1.4). Similarly, one participant asked that the vehicle mute all

non-critical alerts during this task, and that "for this, minute and fifteen seconds [duration of LPD], I'd expect them to run fine on their own. Because I'm busy doing something else... All of [the other systems] would be ignored for this minute and fifteen seconds. If it couldn't be ignored for a minute and fifteen, then it should ring a bell and speak up." (Req. 1.4.2) If an alert was given, then it is the participants' preference to have the warning (auditory and visual) exist until dismissal (Req. 1.4.1). A high-frequency of alerts contributes to workload and several participants reported an increased workload in off-nominal events due to the subsequent assessment of the warning severity (and impact on primary task) and additional decision making to compensate for the warning.

The display suite used in the experiment elicited several design suggestions. Comments and results related to the display suite are numbered 1.1-1.42 in Section 6.4. While the PFD and the LPD displays were used by almost all of the participants, the Overview (OV) display was only used by one participant. Generally, participants reported that the OV provided useful, but non-critical information, as the task did not require any piloting. They also found the representation of the information to be confusing and difficult to monitor. One participant suggested including the terrain profile with the OV, as to illustrate the proximity of the vehicle to the terrain (Req. 1.3.1 and sub-requirements). One participant commented that using the imperial system on the OV display in addition to the metric system was appropriate for future international missions. All other participants requested that the information be presented in all the same units, preferably in metric (Req. 1.3.13). The breadcrumbs and 5-second projection points would be useful for piloting the vehicle or determining if the system was not functioning normally (Req. 1.3.2.1). Unlike the OV, the PFD was used more frequently. Eleven of the thirteen participants polled on display usage reported using at least one element of the PFD. One participant worked solely off the window and did not use the PFD or the OV. Another participant focused only on the LPD display and did not use the other two displays. All participants who used the PFD said the additional camera view provided information critical to landing site decision-making. The time remaining, altitude, and fuel indicators were also used for situation awareness, but did not actively factor in the decision-making process. No participants reported trying to

make a decision by a certain state, such as a fixed altitude, time-to-go, or fuel remaining. Several astronauts did mention wanting to work faster in shorter scenarios. Conversely, some participants used the entire time allocated - 45 or 75s - to make their decision. Less than five participants used the other vehicle state information - pitch, range, and vertical velocity. As with the OV, the participants noted the utility of the information, but the lack of piloting responsibility reduced the significance of these values (Req. 1.3.3-1.3.7, 1.3.14).

Only one participant reported actively and successfully using the window for decision making during LPD. He reported having some initial difficulty correlating all of the images but soon relied strictly on the window and less on the PFD. The window was useful for determining the proximity to nearby hazards and confirming the location of the vehicle with respect to the overall terrain. Other participants attempted to use the window - and many commented on the need for better imaging in the simulation - but had trouble with correlating this information to the other displays. The images and screen resolution presented through the window and the PFD were the same within the experiment. In normal circumstances, the resolution of each would differ, depending on the vehicle altitude (and other occluding effects such as shadowing or dust) and the quality of the camera. There was no particular negative or positive consequence with using or not using the window, and its inclusion was primarily to survey the window desirability, usage, and for simulation representation completeness.

The PFD, the LPD, and the window all provided mechanisms for viewing the terrain. Twelve of the thirteen responses discussed difficulty with correlating all of these perspectives. During debriefing, participants frequently commented on the inability to correlate these images, often ignoring the window (most frequently occurring), one aspect, or the entire display suite. As one participant stated, "we will find what's most important to that which gives us the best data at the time... If we had a really big screen TV, sitting right here, with really accurate [information], that would be the most important [source of data], and you could conclude, oh, they really need that stuff. Or, you could have a cockpit, with a huge window, and no automation and that would be our primary tool of where we get our information from. And so you come out of that thing going, wow, they really need a big

representation of a window, or a presentation from a camera onto a screen, or something that gives you that." Participants reported needing this visibility and correlation as to determine whether the vehicle was landing in the correct area, in addition to determining the validity of the LIDAR results. One participant requested information aid in alignment, such as identifying the same crater geometry across all of the displays (Req. 1.3.2, 1.3.8). Several suggestions were made regarding the use of a Heads-Up Display (HUD) in combination with the window, either supplementing or replacing the PFD. In general, none of the participants reported missing any particular piece of vehicle state information within the bounds of the designated crew role. Several pilots requested more details on the specific cause of the warning, such as a diagnosis. Should the crew's role change - such as the addition of flying the vehicle - more information must be carefully incorporated into the displays.

6.2 Method of representation

Comments regarding the symbolism and location of information presented on the display were frequently made throughout the experiment. Requirements 1.3-1.3.14 reflect the findings of this work. In general, participants suggested presenting the information in an intuitive, visual manner that required less interpretation within the context of the scenario. For example, the consumption of resources such as time, altitude, fuel, vertical speed, and range should be represented as tapes or gauges (Req. 1.3.6). Both methods visually show consumption relative to resource availability, thereby eliminating the step of interpreting the significance of a specific state value. The tape or gauge should also change color if the vehicle is near or violating a threshold (Req. 1.4.1). To further emphasize the significance of this state, the cockpit should provide an audible tone (Req. 1.4.2). The LIDAR scan should be shaded in a similar fashion, with stoplight colors - green, yellow, red - to denote the quality of the terrain (Req. 1.1.3). Red highlighting represents absolute hazards - the lander cannot land here under any circumstances. Yellow highlighting represents terrain that is difficult to land in, perhaps requiring additional orientation or maneuvering to avoid rocks or smaller craters. Green highlighting denotes terrain that is clear for landing, with

no additional maneuvering required. In situations of particularly hazardous terrain, the participants recommended using only the green highlighting. These terrain annotations generally improve overall situation awareness, as noted by Kramer et al. [102] [103], who examined the use of enhanced vision or synthetic vision during the lunar landing approach. Their experiment also featured one-color hazard highlighting, but they did not report a need for multiple colors.

Participants also suggested retaining two symbols (the vehicle, the point of interest) through all of the displays - OV, PFD, HUD - to illustrate the position of the vehicle relative to the intended landing site and to support situation awareness (Req. 1.3.1, 1.3.9). Maintaining visual momentum [104] should be a key focus of the cockpit designer, and ensuring that transitions between displays do not significantly impact the cognitive processes of the cockpit. The intending landing spot should also be identified at all times, whether as a candidate of three automation suggested sites, or supplied by the participant. The color of the circle should change with respect to the task status - one color for a designated site, another base color for under consideration (Req. 1.3.10). To assist with vehicle orientation and terrain alignment, one participant recommended outlining critical crater patterns on each display (Req. 1.3.8).

6.3 Cockpit layout

The arrangement of the displays, the windows, and the use of a window, camera, HUD, or standard displays were frequently critiqued by the participants. Generally, participants preferred having the PFD in the center, with the LPD display to the side. A few commented on the ability to complete the LPD task on the PFD alone, without the use of the LPD display. The primary display should be held at eye level and be perpendicular to the floor (Req. 1.2.5 and sub-requirement). The window - if employed - should be aligned with the displays and be placed in front of the pilot, with emphasis on placing the window directly in front of the pilot. The suggestion was made to angle all of the displays if the window could not be easily reoriented within the design of the cockpit. One participant also suggested tilting the pilot to ensure correct orientation of the window with respect to the person:

"You can tilt me in space, for example, by having me stand one foot on a plate". Both participants disliked the side windows, commenting on the inability to see the terrain at that angle. The arching terrain of the chosen location for the experiment also made one side window ineffective, as the pilot could not see the horizon from that perspective. Four participants suggested the use of HUD, to supplement the window visibility. One pilot asked for a larger window, suggesting that no seats be included in the cockpit design, to allow for closer viewing through the window.

The preference between windows and cameras was mixed, with several pilots asking for a dual system. Participants discussed the merits and demerits of both systems. There were three major arguments made for camera usage: flexibility with placement; visibility capability; and affordability. First, cameras are flexible as they could be placed anywhere on the lander and the resulting image would show up on a display to the crew members. Camera performance is not as tightly dependent on the vehicle's movement as using a window. Unusual vantage points can be afforded with cameras. Likewise, multiple camera systems could be employed to provide stereometric imaging or a wide angle perspective. All live feeds can be processed and annotated before presenting the information to the pilot. Camera systems can improve upon their viewing angle by having a gimbal mount. Instead of statically placing several camera systems to achieve constant visibility, especially as the vehicle approaches the surface, a singular gimballing camera system can be used. The gimbals can be set to automatically target and track a specific surface feature, or the pilot can control the camera itself. While the later option provides a greater degree of control particularly in off-nominal scenarios, participants in the experiment pointed out the additional task involved with having to identify and to set the most appropriate camera angle. Second, cameras can extend beyond the normal viewing range of a human and increase the perception capability. Pilots can receive critical information sooner, thereby having extra time to process information and decide on the final landing site. Lastly, camera systems are reasonably affordable [105, 106] and are commonly in extreme conditions, including surveillance, medicinal, and search and rescue [107, 108, 109]. However, there are disadvantages associated with cameras: adaptation to the space environment, failure

rate, depth imaging. The use of electronic sensors in space requires radiation and thermal shielding, which can be costly in terms of money and mass. Additional costs may arise from employing high-resolution cameras. Like all electrical systems, there is always the risk for equipment failure. While ground and flight testing can minimize and improve the risk for camera failure, the possibility still exists. Designers must plan for likely failure rates and provide methods to compensate for partial or full failure. The last argument against sole dependency on a camera system is the ability of the camera to replace the human eye. As all casual camera users have experienced, the image perceived by the camera can be skewed or lack depth. The perceptibility of the lunar terrain is sensitive to lighting conditions. Naturally occurring lighting angles are capable of minimizing or eliminating depth perception of the terrain. Camera systems have the potential to misrepresent terrain data.

The use of a window has been a topic of interest since the Apollo mission [55] [54] due to the glass density. Early designs promoted large windows, but were quickly reduced to save weight. The focus was on window design efficiency - maximum operator performance for minimum mass and volume - rather than the modern debate of window usage. The improvements in camera technology have countered the continued reliance on windows for anything but backup systems. However, the participants believed that the window has three performance advantages: Negligible risk for hardware failure; affordability; accurate visual representation. Windows provide negligible risk for hardware failure, as they require no power. They may not be as effective if they are obscured or manipulated in some fashion, such as thermal or mechanical stresses, impacts from micrometeoroids, dust occlusion [110, 111]. A cracked or broken window poses the same hazards as a breach in the vehicle's outer hull. Likewise, because windows are not a digital system, they require chemical coatings to protect from thermal and radiative elements as compared to multiple high-resolution camera systems each requiring physical shields and performance benchmarking. Lastly, a window provides depth imaging, as only a few centimeters of glass stands between the human and the terrain. While terrain sensitivity to lighting is a problem that plagues depth perception when humans view out the window, generally depth perception is more readily achievable with human eyes. Despite these advantages, windows have three major

shortcomings: required mass, subsystem dependency, performance limitations. Windows, regardless of the material used, are dense and the achievable visibility is proportional to the volume. Larger mass requires more fuel to move the system, leading to reduced payload capability and increase in mission costs. Additionally, the use of windows and the size of such window creates a substantial subsystem dependency. A large window impacts the size and structural composition of the lander, from fuel tank sizing to vehicle asymmetry to account for such a large center of gravity offset. The vehicle trajectory must be designed to account for landing site visibility, possibly resulting in non-optimal fuel consumption movements. This particular argument was strongly debated during the Apollo mission, as the window was the only source of landing site hazard detection and avoidance imaging [48]. Lastly, windows cannot substantially improve upon the resolution capacity of the human eye without additional strain. Subsequently, data flow is limited to the ability of the human eye to see and interpret the perceived image.

Using a basic relationship regarding visual acuity and spatial resolution, one can determine an initial approximation for the savings in time provided by the LIDAR sensor, a camera system, and a human looking through the window. Visual acuity is measured as the distance required for the eye to perceive an object of 5 arcminutes as compared to standard distances. Generally, this proportionality is given in terms of a standard distance of 20 ft or 6 m, thus, "20/20" or "6/6" vision. However, the geometry of this definition can be extended to the lunar landing scenario for a first order comparison of human perception and sensor spatial resolution. The standard viewing distance in the visual acuity equation is analogous to the distance between the pilot and the hazard. The object is a 30 cm hazard (assuming equal width and height). Therefore, the pilot must be within 206 m of this hazard to see it. This same hazard, however, is detected by the LIDAR scan at the beginning of the LPD maneuver, at a range of 866 m. Subsequently, information regarding hazard location is received 35s sooner than viewing through a window. The same calculations can be made through a camera system and compared with this approximation of human performance to provide a measure of how soon information can be received [112].

Evaluation of analogous systems illustrate a balance between cameras and windows. The

lunar Altair lander, part of the canceled Constellation project, proposed the use of a two cameras and a window. The two cameras - narrow angle and wide angle - would be mounted on a gimbal to provide optical navigation from lunar orbit insertion to touchdown. Riedel et al [113] proposed the Mars Reconnaissance Orbiter's Optical Navigation camera to provide a high-resolution, narrow angle and the Charles Stark Draper Laboratory Inertial Stellar Compass (ISC) as the lower resolution, wide angle camera system. It is assumed that future landing systems would consider both cameras as a baseline for optical navigation and supporting crew decision making. At the time of cancellation, the Altair team was conducting studies on window geometry and placement [88]. The Apollo Lunar Excursion Module used three windows - two side windows and a docking window.

Regardless of the implemented system, it is worth considering spatial disorientation effects during lunar landing. Clark et al. [46] developed a model of disorientation and compared the changes in perceived orientation between the Altair and Apollo landers. While both effects are small, the Altair lander caused a greater disorientation effect than the Apollo lander.

The discussion between the HUD and standard displays is less heavily debated compared to cameras and windows, as the focus is more on information layout rather than the method of information sensing. HUDs are often suggested for use with windows, to replace the role of camera systems or standard displays. The premise behind these systems is to support divided attention by improving the spatial proximity of the user and the external environment. Other cockpit display systems have been suggested for use during the lunar landing scenario. The participants were not questioned about these systems specifically, nor did any mention them. Arthur et al. [114] has suggested the use of a Head-Worn Display (HWD) in lieu of a HUD. He states that unlike the HUD, a HWD has an unlimited field-of-view, is lighter, and is not constrained to be forward-facing. Overall, the HWD improved situation awareness compared to Heads-Down displays (traditional), but additional training is needed to improve confidence with operating this new system. Similarly, Highway-in-the-Sky, a means of projecting the expected future vehicle state was proposed as another concept. Forest et al. [76] solicited pilot opinions on this flight path display and

did not find an agreement for or against.

6.4 System requirements to support Apollo-like and moderate function allocations

LIDAR

1.1 The System shall conduct and provide the results of multiple LIDAR scans with no change in resolution.

1.1.1 The System shall conduct and provide the results of LIDAR scans at 0.2 m/pixel resolution of the designated area of 180m $\times$ 180m and at 0.25 of the lander pad diameter per pixel for zoomed in scans.

1.1.2 The System shall provide the Pilot the option to remove the LIDAR scan results from the displays and the Visuals.

1.1.3 The System shall identify two levels of hazardous landing areas: No landing, possible landing.

- A hazardous level of No Landing shall be defined as: A vehicle footprint that does not have any clear areas for the lander pads and the main engine.
- A hazardous level of Possible Landing shall be defined as: A vehicle footprint that contains enough clear area - but may still contain hazards - to support the lander pads and the main engine. The vehicle may land there but additional adjustments are necessary.
- A safe landing spot shall be defined as: A vehicle footprint that contains enough clear areas to support the lander pads and the main engine without additional adjustments.

1.1.4 The System shall provide the Pilot the option to highlight only the good landing areas.

Visuals

1.2 The System shall provide a real-time, high resolution Visual of the expected landing area.

1.2.1 The System shall provide a continuous visual of the intended landing site until touchdown.

1.2.1.1 The System shall provide a visual of the landing pad placement relative to the terrain.

1.2.2 The System shall provide external lighting such that sufficient contrast (???) of the landing area is maintained

Visual Performance

1.2.3 This Visual must provide a resolution of XX m/px or better at a rate of YY Hz. The image contrast must be naturally occurring (sun and vehicle approach angle geometry) or artificially provided (external lighting source).

1.2.3.1 The real-time high contrasting Visual supplied by the System must include Vehicle state information, suggested landing sites, and hazard identification.

1.2.4 Overlays on all Visuals shall provide sufficient contrast for the Pilot.

Visual Placement

1.2.5 The System shall be arranged as such that the visuals are perpendicular to the lander floor.

1.2.5.1 The System shall be arranged as such that visuals are placed in front of the Pilot, with the intended Primary visual mounted at eye level.

Critical Information

1.3 The landing System shall provide task-critical information to the Pilot and the crew relevant to Landing Point Designation.

1.3.1 The System shall provide information regarding the current position of the Vehicle and location along the expected Trajectory.

1.3.1.1 The System shall provide a visual and audible warnings if the Vehicle is within 10s of colliding with the terrain.

1.3.2 The System shall identify the vehicle location with respect to any visual of the landing area.

1.3.2.1 Immediately after the LIDAR scan, the System shall provide information regarding the current (t = 0s) and the expected (t = +5s, t = +10s) vehicle altitudes and proximity to terrain.

1.3.2.2 Immediately after the pitching maneuver, the System shall project terrain alignment guides on all visuals of the cockpit until such information is dismissed by the Pilot.

1.3.3 The System shall provide information regarding the current orientation of the Vehicle.

1.3.4 The System shall provide information regarding the current velocity of the Vehicle.

1.3.5 The System shall provide information regarding the time remaining for Landing Point Designation and the time until Vehicle touchdown.

1.3.6 The System shall provide information regarding resource consumption, where the current value is represented as a proportionality of the total available for that resource.

1.3.7 The System shall arrange information on position, orientation, velocity, and scenario timing on the primary display.

1.3.8 The System shall provide terrain alignment guides on all visuals of the expected landing area, to assist in Pilot orientation.

1.3.9 The System shall identify the point of interest along all of the visuals through all times during the Landing Point Designation task.

1.3.10 The System shall visually identify whether the landing site has been designated or is currently under consideration.

1.3.11 The System shall identify the automation's ranking of all suggested landing sites.

1.3.12 The System shall provide landing site scores of the designated or suggested landing sites.

1.3.13 The System shall provide all information in SI units.

1.3.14 The System shall provide information regarding the Vehicle health and resources but during Landing Point Designation, this information will not be present on the primary display.

Warnings

1.4 The System shall notify the Pilot when nearing or violating threshold values.

1.4.1 The System shall alert visually and audibly, but not require acknowledgment, of any non-critical system failure or malfunction during landing point designation and terminal descent.

1.4.1.1 A non-critical system failure is defined as failure of a subsystem that does not require immediate attention from the Pilot and can be addressed at least two minutes after activation without significantly impacting the completion of the Landing Point Designation task.

1.4.2 The System shall provide a visual and audible warning if the Vehicle is approaching a critical threshold, unless the threshold violation is non-critical to Landing Point Designation.

6.5 Proposed Function Allocation

Chapter 3 introduced a decomposition of the LPD task into five fundamental elements: sensing, interpreting, creating, selecting, and executing. The moderate and Apollo-like function allocations were defined based on a distribution of these tasks to the automation, the human, or shared between both. Throughout the experiment, participants commented on the desired role of the automation. This question was not directly posed and only three participants made direct statements referring to the automation capability. However, this information, in addition to an analysis of other comments and design suggestions, has painted a picture of the necessary automation functionality to support astronaut decision making.

Several of these five fundamental tasks were observed to lack sufficient granularity and thereby can be decomposed to into subtasks. The task of collection information can be performed by both the human and the automation sampling directly from the environment (e.g., human perceiving the terrain, automation using LIDAR scans), or at a low level. The human and automation may collect information at a meta, or high, level, by sampling aspects of all collected information (e.g., human chooses only to focus on the altitude but automation supplies all data). The interpreting task can also be decomposed into comparison and significance. At comparison, the automation and the human identify hazards and remaining resources by predefined safety thresholds. Significance, however, occurs in addition to comparison as the human identifies the impact of such information to the goal. A piece of information that violates thresholds, but is not important to the goal, is effectively

irrelevant. If the conditions are nominal, then the goals of the human and the automation are the same. However, in off-nominal scenarios, the goal has changed and the automation's pre-built definitions and states may no longer be relevant. The tasks of creating, selecting, and executing decision options comes in two forms: coarse and fine, which mimics the areal/local search method used by the Apollo-like pilots. At the coarse level, the human or the automation creates options and selects as to where is the best spot to place the lander. After this selection, the human or automation flies to this landing site. At the fine level, the human or the automation creates options and selects the specific *orientation* or nearby landing site best suited for the vehicle. Execution of this decision selection focuses on placing lander pads between smaller rocks and craters. However, as early as 20m altitude [115], this capability may not be possible due to the dust plume caused by the engine. Both human and automation systems would have trouble with low altitude adjustments as visibility is effectively eliminated.

Table 26: **LPD Task Decomposition and Function Allocation within Modified Endsley-Kaber framework.** Entries marked with [1] indicate that neither the automation or the human controlled the Vehicle orientation, which was fixed to windows-forward, perpendicular to the landing surface.

	Sensing		**Interpreting**		**Creating**		**Selecting**		**Executing**	
Allocation	**Hi.**	**Low**	**Com.**	**Sig.**	**Co.**	**Fi.**	**Co.**	**Fi.**	**Co.**	**Fi.**
All Crew	H	H	H	H	H	H	H	H	H	H
Apollo-like	H	A/H	H	A	H	H	H	A[1]	A	A
Moderate	H	A/H	H	A	A	A	H	A[1]	A	A
Robotic	A	A	A	A	A	A	A	A	A	A
Desired Nom.	A/H	A	H	A	A/H	H	A	H	A	H
Desired Off-Nom.	A/H	A	H	A/H	A/H	H	A/H	H	A/H	H

With this new framework and the responses of the automation, Section 6.6 is rewritten to include the desired responsibilities of the human and the automation. The desired nominal function allocation follows the distribution of tasks as prescribed in Table 26. In this function allocation, both the sensing and interpreting tasks are completed by the automation as both entities share the same goal. In creating, selecting, and executing, the automation handles the high level task of determining where to place the lander (Req. 2.1). The human is in charge of determining the vehicle orientation that best accounts for the

rock and crater distribution (Req. 2.2.1, 2.2.2). One participant stated his belief that orienting and correcting the vehicle over small distances best leverages the pilot's ability, whereas determining where to best place the vehicle is the job of the automation. The automation has more processing capability to quickly and accurately collect data and make one optimization. If there are multiple "equally best" options (see Chapter 5 for an example of equal sites when determining a landing area), then these options can be presented to the pilot for a human-determined tiebreaker (Req. 2.1.1). For the final adjustments, the pilot can correct much faster than the automation as the vehicle approaches the surface. Until the automation can complete the cycle of collecting real-time data, identifying localized hazards, creating and selecting an appropriate orientation or position change, and then executing the necessary commands, then the human is the best agent for the job. Although no approximation currently exists for this particular task, one can estimate the timing of the processing cycle based on related work.

Assuming nominal operations and a trained astronaut, the LPD task simplifies to a set of rapidly occurring actions: eye movement time, visual processing, cognitive processing, and motor processing. Based on the estimations derived from the Model Human Processor [24], one can approximate that the human completes this processing cycle within 175-1170 ms, or 1.17s. This value exceeds the present approximation of 5s for engaging the LIDAR sensor, processing the results, and selecting an optimal solution. Therefore, until automation can perform this task in less than 1.17s, the human complete the task of placing the landing feet more efficiently than the automation. However, a caveat exists with this approximation - within 20m altitude, dust plumes due to the engine proximity to the ground causes visibility issues.

The landing point designation task should be simplified as much as possible for the pilot, as to reduce the number of options to evaluate and choose between. The trajectory should be designed to minimize the importance of saving fuel by landing near the POI or landing quickly, so the pilot does not feel compelled to account for fuel consumption in his decision making process (Req. 2.3 and subquirements). The guidance, navigation, and control system used by the Autonomous Landing and Hazard Avoidance Technology

(ALHAT) team was designed to reduce the significance of fuel savings. Manuse et al. [116] noted that a pilot could achieve a divert to the edges of a 360×360m landing area, diverting twice as far as the LIDAR scan ($180x180$m grid) and have sufficient fuel. Additionally, a map of the delta-v required to divert shows that within the areas of the LIDAR scan, the difference is 2 m/s. If training matches expectations, then the process to finalize the landing location will be routine, a confirmation of the correct state. However, if training differs from expectations, then the pilot must realign his mental model to the reality of the situation, interpret the effect of the discrepancies between his mental model and reality, and then execute the necessary commands to reach his goal. As seen from the experiment results (discussed in Chapter 4), the astronauts complete this task between 15-25s, with many pilots purposefully taking the full allotted time. A similar result has been seen in other high-stakes, time-critical scenarios such as firefighting [117]. Since the participants "will find what's most important to that which gives us the best data at the time" and they are expecting updates on terrain and vehicle state as the task continues, it is reasonable to assume that they will try to maintain control during the entire landing sequence (Req. 2.2.1).

The desired off-nominal function allocation distributes more of the workload to the human, as the priorities of the human may have shifted. Depending on the scenario, the human may opt to discard secondary goals or aim to reach the ground faster. Subsequently, the human must have the control authority to adjust the information on the displays, and to accommodate for any useful information the automation is able to provide. For example, even within the LIDAR warning scenario modeled in the experiment, several participants reported using the faulty LIDAR data, as it might contain valuable information. They stated that there was some trust in the LIDAR scan, as some of the information matched their perception and self-identification of the terrain. In the case of creating, selecting, and executing, the human must now make critical decisions with this imperfect data. Aside from having the option to declutter the display, the pilot should still have the opportunity to view the raw data (e.g., the LIDAR scan with slope and vertical protrusions identified, but no post-processing to account for lander pad and engine bell placement) (Req. 3.2.1).

In an off-nominal scenario, the pilot has the additional task of redefining the new landing objective and what information is necessary to support this goal. As stated by the participants, the increase in workload comes from diagnosis and filtering out non-useful data. The automation works best by assisting the astronaut in this diagnosis. Similar to all-human teams, the combined human-automation system will act efficiently if both members share the same mental model. The automation must be able to anticipate the pilot's actions, suppress excessive or unnecessary data, and support situation awareness. For example, a simple mechanism to minimize workload during such situations is suppressing bad LIDAR data, a feature most requested by the participants. Another example of supportive automation during off-nominal scenarios is the engine warning scenario presented in the experiment. Many participants expressed a desire to land quickly, intending to landing short of the POI or following "the sort of rule in aviation [that] if you're having trouble, don't fly over a good landing site". A supportive automation system may suggest a "quick landing site" option, that minimizes the amount of time required to safely place the vehicle on the ground. Or, if the automation is aware that the engine is not producing as much thrust, the automation may remind the pilot that the vehicle cannot decelerate as quickly as before and the easiest site to land at is downrange of the POI (Req. 3.2). To support situation awareness of the severity of the incident, the automation can switch camera viewpoints to the part of the vehicle concerned. To facilitate and minimize bottlenecks of the pilot's sensory channels, the automation may consider using natural language [118, 119, 120, 121, 122].

Subsequently, a new set of system requirements can be written to account for this desired functionality of future automation systems. Section 6.6 accounts for these new system requirements, specifically for the creating, selecting, and executing tasks. These requirements complement those presented in Section 6.4.

6.6 System requirements to support ideal nominal and off-nominal function allocations

Automation Role, Nominal

2 The Automation shall support the Pilot through nominal operations

2.1 The Automation shall suggest an optimal landing site after each LIDAR scan, based

on mission criteria.

2.1.1 The Pilot shall be given control authority to select between multiple optimal sites.

2.2 The Automation shall control the vehicle to the intended landing site, unless the Pilot takes command of the Vehicle by physically engaging the stick or redesignating the landing site.

2.2.1 The Pilot shall have the authority to control the vehicle and make corrections to the final landing site location within the landing area.

2.2.2 The Pilot shall have the authority to control the vehicle orientation.

2.3 The Vehicle should be designed such that there is sufficient fuel for the LPD task.

2.3.1 The Vehicle trajectory should be designed such that sufficient fuel exists to land at any point of the LIDAR scan.

2.3.2 The Vehicle trajectory should be designed such that sufficient fuel exists to land late within the designated LPD time, rendering consideration of the fuel consumption negligible.

Automation Role, Off-Nominal

3 The Automation shall support the Pilot through off-nominal operations.

3.1 The Automation shall visually and audibly alert the Pilot if a critical malfunction has occurred.

3.1.1 The visual and auditory alert shall persist until the Pilot has dismissed the alert.

3.2 The Automation shall diagnose the critical malfunction when thresholds have been violated and suggest possible solutions.

3.2.1 The Automation shall present suggestions to compensate for a critical malfunction to the Pilot through visual means.

3.3 The Pilot shall override the Automation at any point in the landing sequence.

An example of this cockpit layout, incorporating the system requirements from this Chapter, is presented in Figure 41.

6.7 Training

The experiment debriefing illuminated a number of biases and assumptions made by the participants, many of which can be eliminated by designing a system to compensate for these biases, or through training. Analysis of participant feedback illustrated two main areas of confusion: incomplete understanding of the terrain and a misunderstanding of the vehicle and automation's capabilities and limitations. While only half studied the photograph of the expected landing area, all pilots, in some manner, wanted to know more information regarding the expected landing area. As discussed in Chapter 4, it is shown that mental modeling reduces the amount of additional work the pilot must complete during the task itself. Developing a mental model of the expected terrain provides a nominal solution for sensor data. And, as several participants noted during the debriefing, knowledge of the terrain can reduce the impact of a bias. Six participants commented on being uncomfortable landing in areas of complete darkness, factoring this criterion into their decision making. One participant acknowledge that "[landing in the dark spots is] an action I know I would take if I know that I am highly trained and the sensors are reliable to do so and I'm able to cross-check it. If that is what the mission calls for, and all safety guys say we can do this, then that's what I'm going to go for, that's what I would truly do if I had the training". Four participants also commented that having better knowledge of the terrain would assist with cross-checking the performance of the vehicle navigation. A known crater formation can be used to re-orient oneself to the surface. Trusting the vehicle navigation and being oriented to the terrain also improves situation awareness. One participant also discussed creating 3-D images in her head of the terrain. Such elevation maps would be of use to astronauts during training. Lastly, as navigational errors are possible, astronauts should learn about the terrain around the intended landing site, within the expected landing footprint due to such dispersions in initial vehicle state. It is not possible for the astronauts to know each aspect of the terrain, thereby increasing the need for a robust automation support system and complete knowledge of the system's capabilities and limitations.

The human-automation system acts as a unified team, requiring knowledge of each entities capabilities and limitations. Beyond this laundry list, the human must understand the

internally programmed goals and biases of the automation. Threshold values such as "hazardous" must be clearly explained, just as the automation must be provided circumstances of which craters and when can some be safely ignored. This particular area of confusion was highlighted several times by the participants: "it may exceed the lander tolerances...". Additionally, the development of the scoring algorithm (Chapter 3) and the cognitive model (Chapter 5) proved to be instrumental in determining what logic must be made transparent to the pilot. Often, results from the cognitive model or the scoring algorithm would be unclear in terms of the criteria employed, definitions coded, and relative importance assigned. If this information is not clear to the operator, or a justification of differences between expected and actual results, the pilot is unlikely to trust the automation. Likewise, mission preferences versus mission capabilities must be clear to the astronaut, as to the consequences of violating thresholds. One of the most common questions posed by the participants was, "What am I supposed to look at? What is most important here? What are my main priorities?". After the automation role has been defined, astronauts must be trained on his specific roles and responsibilities. He needs to comprehend the priorities of the mission and what attributes can be discarded if necessary.

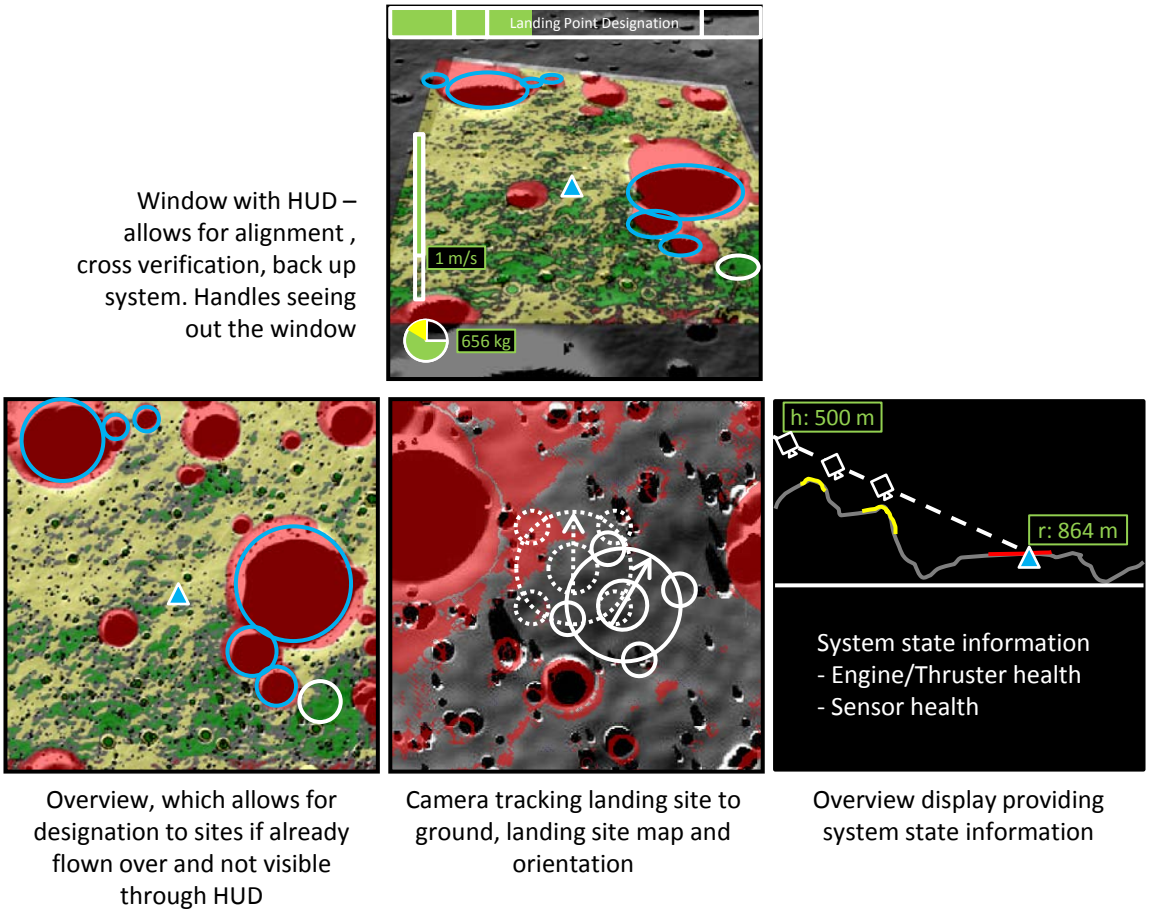


Figure 41: **Notional cockpit layout for desired function allocation.** This display arrangement consists of three displays and a window with a HUD. The window provides another method for surveying the landing site and includes digital annotations to assist with orientation and decision making. The bottom displays, from left to right: a still frame of the expected area including degree of hazard highlights and orientation aids, with continuous updates on expected landing site and location of POI; a camera feed of the expected landing site including expected and intended positioning of the vehicle's engine and lander pads; a display presenting system state information and a trajectory profile including terrain proximity warnings and location of point of interest.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

A study was conducted to determine the functionality for an automated system designed to assist astronauts during lunar landing point designation. Understanding the functions and decision making of humans helps in determining the roles of automation systems that better leverage the strengths of both agents (human, automation) while improving overall performance. Such a collaborative relationship is critical when landing in high-risk, time-critical scenarios, like the south pole of the Moon, Mars, and nearby asteroids. The following research question was posed: How should the functionality of automated landing systems and crew be assigned to support mission operations during landing point designation? And equally important, how does landing point designation performance vary in each of these function allocations?

To address this question, a cognitive process model was developed based on empirical data of astronaut decision making for use in requirements analysis. The cognitive process model is based on current literature and designed to reflect the findings of several human-in-the-loop experiments. The overall research task was completed in three research endeavors and offers the following three major contributions, as proposed in Chapter 1:

1. Characterize human-system decision making during landing point designation and provide a quantitative database of this performance.

2. Develop a cognitive process model to establish performance benchmarks and expected achievements during conceptual design

3. Provide system requirements regarding information needs and cockpit design requirements for humans and automated systems during landing point designation.

7.1 Characterizing astronaut decision making

The behavior of astronaut decision making within different function allocations was characterized through two human-in-the-loop evaluations. The first experiment was conducted with fifteen helicopter pilots and focused on identifying the different decision making cues and search methods used by pilots during the landing point designation scenario. The second experiment, performed with fourteen members of the NASA Astronaut Office, examined the changes in decision making behavior through various function allocations, landing trajectory profiles, and a comparison of off-nominal and nominal operations. Both experiments highlighted decision making strategies and performance differences in the landing point designation task. The experiment results also provided initial approximations for time to complete, landing site safety assessment, and expected changes in decision making. The first experiment with helicopter pilots was conducted to establish the range of cues, search methods, and techniques to incorporate empirical data into cognitive modeling. The cognitive process model designed and simulated in this thesis is based on data collected solely in the second experiment.

Astronauts complete landing point designation in different ways, depending on the function allocation. At a moderate function allocation, astronauts use re-ranking or eliminating search methods to differentiate and select between three potential landing sites. Their focus is on the terrain within and immediately outside of the landing footprint. At an Apollo-like function allocation, astronauts use areal or local search methods to create and select suitable landing sites. They may opt to narrow their selection space in two rounds - focusing on finding a good region, or a subset of the landing area, and then a good site within that good region. These astronauts are focused less on the terrain within and immediately outside of the landing footprint, but are more concerned with general terrain features, evaluating subsets of the terrain as necessary. Within both function allocations, participants performed the task faster when preparing a mental model prior to the task.

Function allocation of the landing point designation task is of significant importance for future crewed missions to the Moon, Mars, and other destinations. However, very few quantitative information exists regarding timing approximations, quantification of landing

site safety, and decision making strategies (search method, number of cues, which cues, and relative importance). This data provides initial approximations for mission designers, allowing for better approximations of maintaining vehicle states to support astronaut decision making. Furthermore, this work provides an exemplar of the development and use of cognitive models in the overall systems design process.

7.2 Developing a cognitive process model

A computational cognitive process model was developed from the data collected in the main experiment with the Astronaut Office. This descriptive, rule-based cognitive process model presents the spread of potential landing sites chosen by astronauts during the landing point designation task, with search methods, cue selection, relative importance, and task time to complete mirroring trends observed in the human-in-the-loop study. Both types of function allocations and the associated differences in behavior are captured in this model, in addition to the two types of trajectory profiles and performance deviations in off-nominal scenarios. Validation of the cognitive process model was completed by evaluating each experiment run and ensuring that the model chose approximately the same site as the participant, given the same contributing factors. Initial results with the cognitive process model were conducted with 1000 randomly generated model data points. The distribution of these model data illustrate that the model selects sites at approximately the same distribution and spread of the actual participants, but also selected sites not chosen by the participants. These points account for the variability in astronaut decision making not immediately observed with this participant sample.

This cognitive process model accounts for interactions between an astronaut and two forms of automated systems: a moderate function allocation, which designates the responsibility of creating a landing site to the automation and selecting from these sites to the human, and an Apollo-like, where creation and selection are solely from the human. The model also represents the changes in behavior during off-nominal scenarios, and the differences in decision making with long and short trajectory profiles. When given a specific landing spot, function allocation, and an operational scenario, the cognitive process model

predicts where astronauts might choose to land, how many and which criteria and their relative importance that was used to generate the decision, the search method employed to find such a decision, and the time required to complete the overall task. This cognitive process model can be used for further analysis of the complete human-automation interaction during landing point designation.

7.3 Proposing design requirements for automation systems

System requirements were derived from an analysis of a cockpit evaluation conducted during the main experiment, cognitive process model trends, and observations of participant behavior. Participants were asked to provide feedback on the cockpit layout, the display representation, and the information needs. These comments were compiled and used to support the development of system requirements for the design of automation functionality. The composite behavior of the participants was also considered and system requirements derived to support behaviors not explicitly mentioned by participants. A normative model of landing point designation was also derived from participant comments and the cognitive process model results. In this normative model, the automation handles the majority of the functionality during nominal operations, but grants control of specific functions to the human in off-nominal scenarios.

The feedback from participants, results of the cognitive process model, and observations of overall human-in-the-loop behavior illuminated a number of information needs, display representations, and cockpit layouts. System requirements for automation functionality were derived from this data analysis for use during requirements analysis and conceptual design. Furthermore, design considerations for future missions are presented, including training topics to counter cognitive biases and participant confusion observed in the human-in-the-loop study. These design considerations also include a discussion of cockpits and windows and the metrics for additional analysis beyond current literature. Lastly, these system requirements are characterized by a fusion of human factors analysis and traditional systems design. Requirements are written as explicitly as possible, with numerical guidelines for automation benchmarks, and human behavior. A normative model of landing point

designation is also provided, to describe the function allocation most suitable for leveraging human and automation strengths during nominal and off-nominal scenarios.

7.4 Future Work

The thesis work conducted in this area provides a number of new research directions. There are three possible future work directions: improvement on the cognitive process model; predicted landing performance; and information processing analysis. The cognitive process model developed for this thesis is a stochastic model based on proportional distributions. As such, the cognitive process model can be improved upon with the inclusion of probability distribution functions that may more accurately match astronaut behavior. Frameworks such as Bayesian inference, neural networks, or Hidden Markov Modeling use these distributions. However, these probabilities cannot be determined from the empirical data collected in this thesis, as the experiment was not designed to solicit that type of data. Additionally, the cognitive process model makes assumptions regarding cognitive phenomena such as perception, memory, learning, and attention. For this thesis, these functions were not needed for the requirements analysis. However, including these functions would improve model fidelity and allow for greater generalization of the inputs to the cognitive process model (e.g., the terrain, the automation functionality). Lastly, the cognitive process model did not focus on the execution of the vehicle or the associated physiological and behavioral functions associated with this subtask. For more detailed analysis, such functionality should be included.

Even without these improvements to the cognitive process model, mission designers can ascertain useful information regarding astronaut performance during lunar landing. The cognitive process model can be used to predict where astronauts would land in both nominal and off-nominal situations. The cognitive process model can be used in tandem with other models - environment, trajectory, vehicle - and the comprehensive system simulated through variations in terrain characteristics, vehicle state, and function allocation. Analysis of these simulated runs can provide trends, cognitive biases, and projected system performance that can be fed back into the design cycle for improvements to the vehicle and automation system.

The last decade has resulted in a significant push to return to the Moon, specifically, the South Pole. Shackleton crater, located at 89.9°S, 0°E has been identified as a region of scientific [123] [124] interest. In particular, NASA has identified a region around the rim of Shackleton that offers a number of operational advantages. This region, at 89.68°S, 166°W, is in an area that is permanently within sunshine, and is lit for 86% of the year ([125],[126]). NASA identified this location as the prime target for the Constellation architecture ([127]). The area is approximately 1 × 5 km wide and could support a variety of missions, from resource collection, observation, and habitation. Likewise, this area is within close proximity to another point of "eternal light". Bussey and colleagues determined that collectively, these two sites are lit 94% of the year, making the tandem an ideal region for habitation and power generation. The cognitive model can provide initial approximations of where astronauts would choose to land, the first set of data for this particular area of future crewed exploration.

The cognitive process model can also be used to perform an initial information processing analysis. The cognitive process model makes decisions based on a set number of cues, which are effectively pieces of information perceived by the astronaut from the displays and the environment. However, as observed in this experiment, the decision quality can be affected by the quality of the information. Since no sensors operate perfectly and redundancy is needed both in operation and in representation, this cognitive process model can be used to determine which information elements require the most design consideration. Such results can improve overall display interface design, assist in decisions regarding sensor redundancy, and anticipate human-automation interaction phenomena such as automation surprise, task workload, and attention tunneling.

APPENDIX A

SUPPLEMENTARY EXPERIMENTAL MATERIALS

This Appendix contains materials used during the debriefing session of the experiment and detailed results from the analysis.

A.1 Experiment debriefing questions

These questions were used for both Apollo-like and moderate automation runs, with slight differences.

First question: "Which statement best matches your overall strategy for choosing a landing site?"

- Apollo-like function allocation
  - "I selected a region based on a set of criteria and then picked a site based on a second set of equal or differing criteria." (Areal)
  - "I eliminated sites based on a set of criteria and then continued until one set was left." (Local)
- Moderate-like function allocation
  - "I eliminated one or two sites based on at least one criterion." (Eliminating)
  - "I re-ranked sites based on one or more criteria and chose the best site of this re-ranking." (Re-ranking)

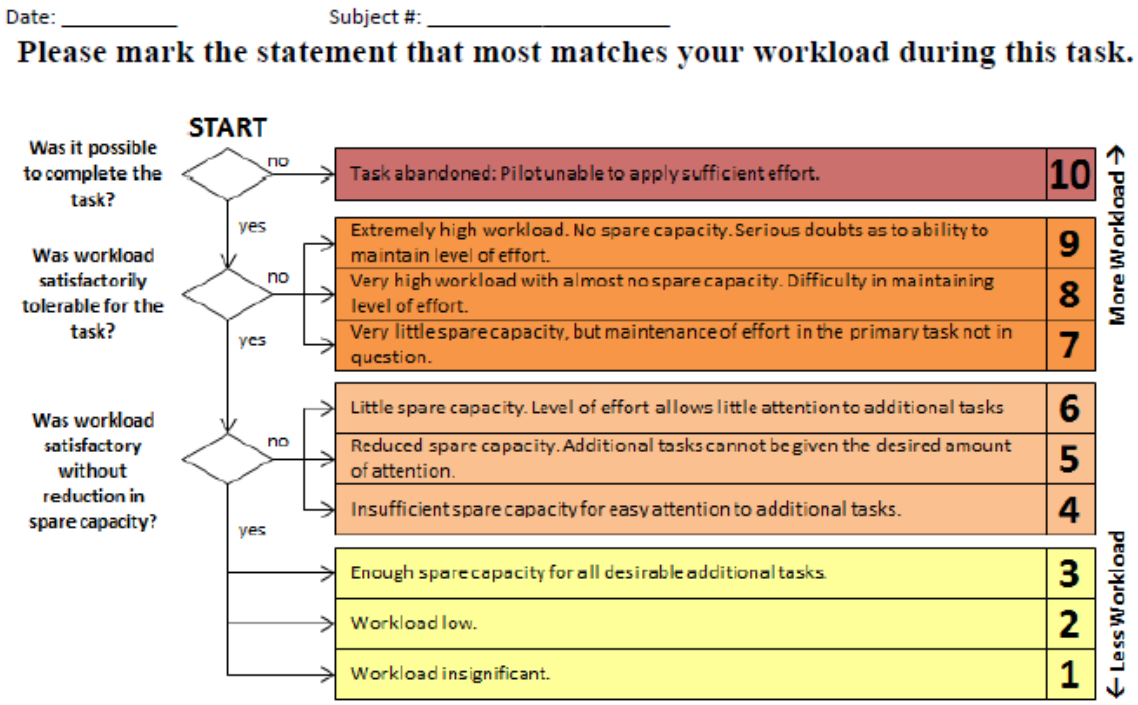


Figure 42: **Bedford scales for assessing workload.**

Date: _____ Subject #: _____

APOLLO-LIKE PRE-EXPERIMENT: Strategy & Landing Site Preferences **0**

"Which statement best matches your overall strategy for choosing a landing site?"

[ ] I selected a **region** based on a set of criteria and then picked a **site** based on a second set

[ ] I eliminated sites based on a set of criteria and continued until one set was left.

"What kind of a site were you looking for?" [ ] no change

Approach

_____ Sites that DIDN'T HAVE/HAD hazards along the approach

(How far along? (D) 0.5 1 1.5 2 2.5 3 3.5)

Interior/Exterior Hazards

_____ Sites that are AWAY/NEAR from large craters *(large is: (D) 0.25 0.5 1 1.5 2 2.5 3 3.5)*

_____ Areas that are least dense with hazards

_____ Sites that DON'T HAVE/HAVE hazards inside the landing area

_____ If placement matters: Hazards shouldn't be on the: *outside inside dead center*

Color/Darkness

_____ Sites that have a uniform color within the landing site:

_____ What color? *White light grey dark grey black*

_____ What do the colors mean to you? *Altitude slope*

_____ Sites that do not have a uniform color within the landing site:

_____ What pattern? *Circular side-to-side top-to-bottom diagonal*

_____ Sites that ARE NOT/ARE in complete darkness

Fuel/Location of sites/Automation Ranking/Time

_____ Sites that are located: *downrange/upperhalf reverse downrange/lowerhalf dead center*
Left right upper left upper right lower left lower right

_____ Sites with a HIGHER/LOWER automation ranking

_____ Perform LPD faster than normal

Used other information

_____ Fuel ____ Altitude ____ Pitch ____ Speed ____ Range ____ Orientation

Expectations

_____ Sites in previously identified (mentally) as good regions

_____ Sites in previously identified (mentally) as bad regions

Figure 43: **Expectations photograph for baseline, nominal.**

Date: _____ Subject #: _____

APOLLO-LIKE: Expectations (AP_B_NO – Fargo)

Please highlight the areas you considered "good" (green highlighter) and "bad" (pink highlighter)

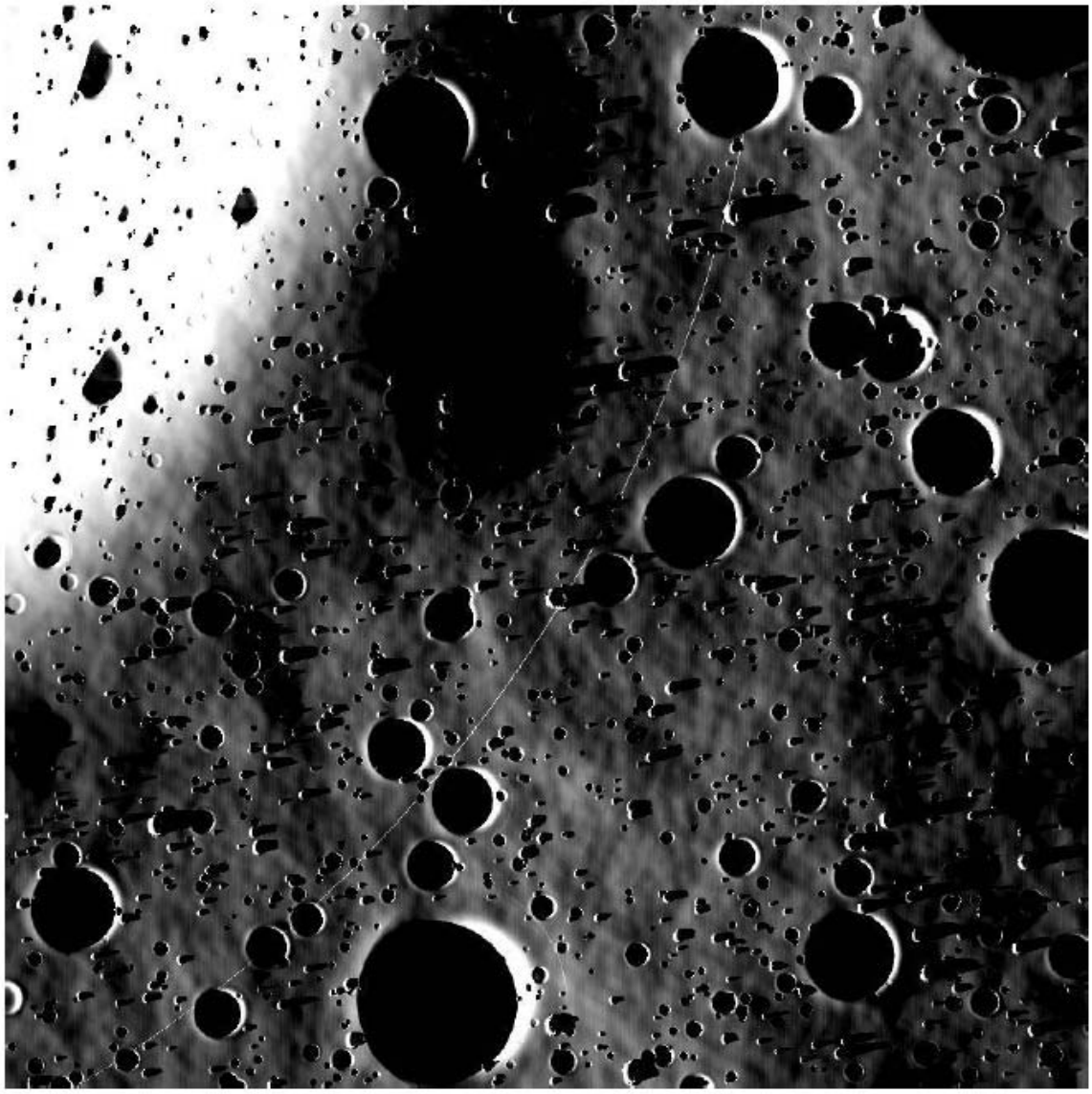


Figure 44: **Expectations photograph for baseline, LIDAR warning.**

Date: _____ Subject #: _____

APOLLO-LIKE: Expectations (AP_B_LI – Tulsa)

Please highlight the areas you considered "good" (green highlighter) and "bad" (pink highlighter)

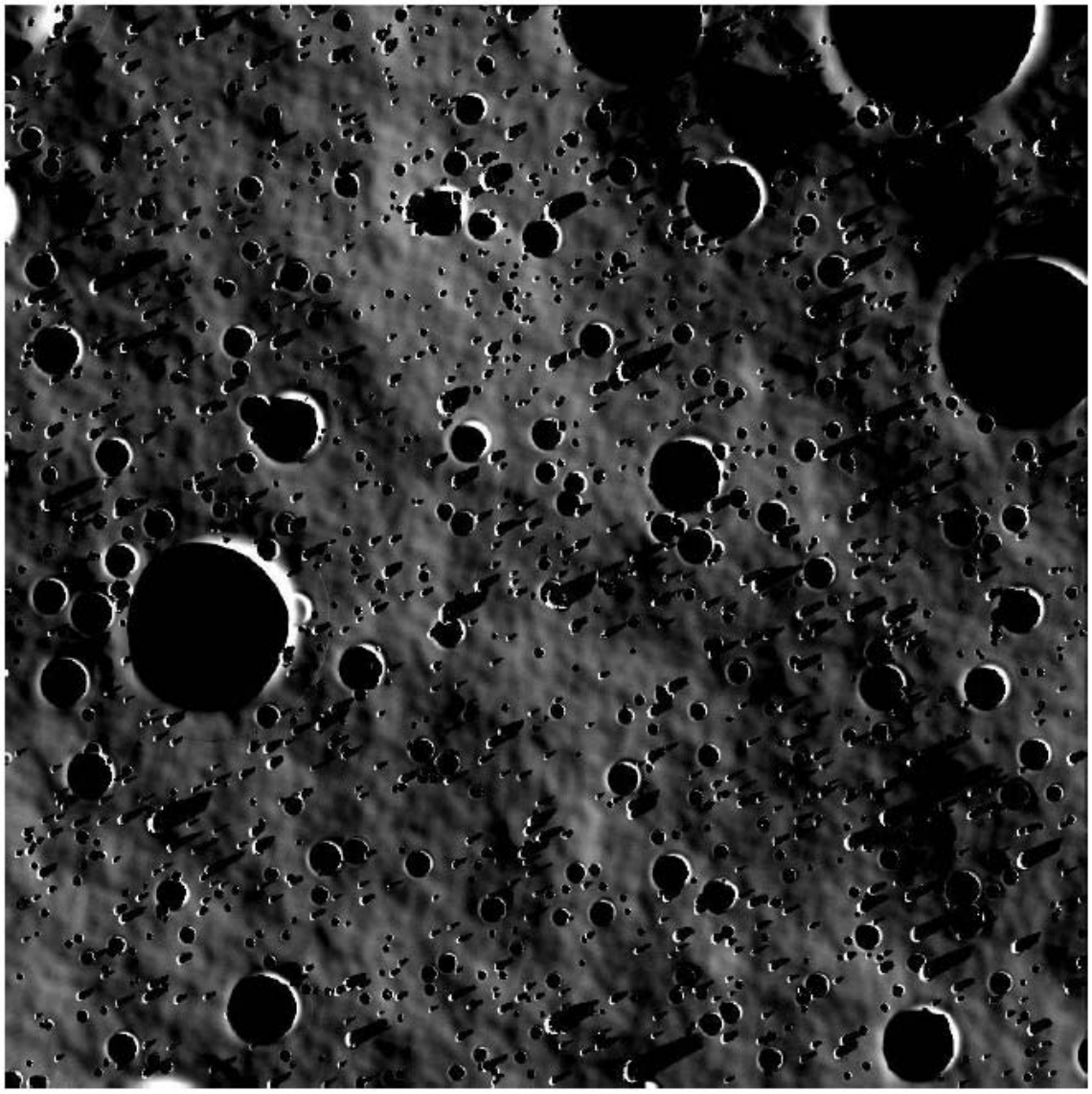


Figure 45: **Expectations photograph for shallow, nominal.**

Date: _____ Subject #: _____

APOLLO-LIKE: Expectations (AP_S_NO – Annapolis)

Please highlight the areas you considered "good" (green highlighter) and "bad" (pink highlighter)

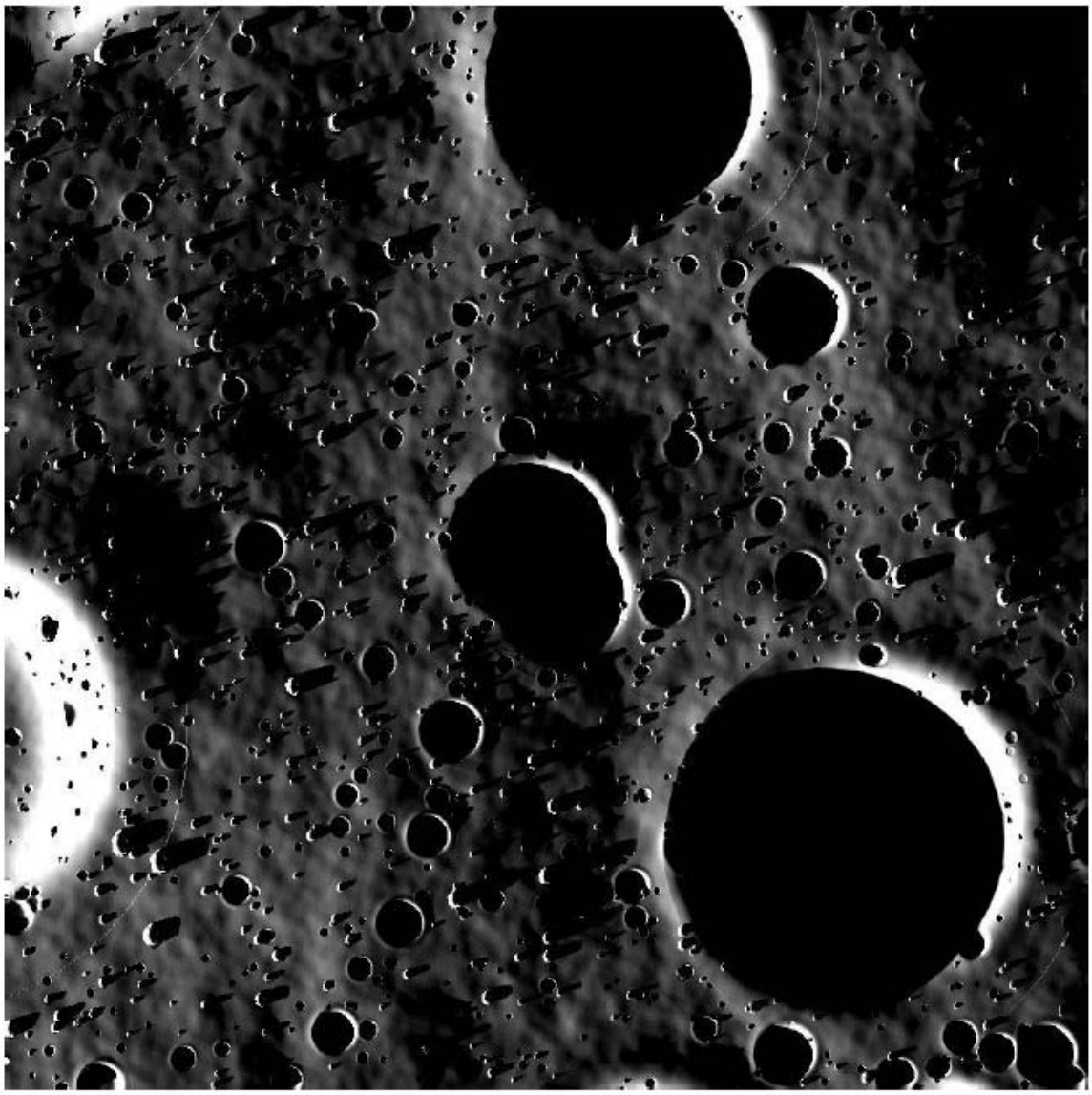
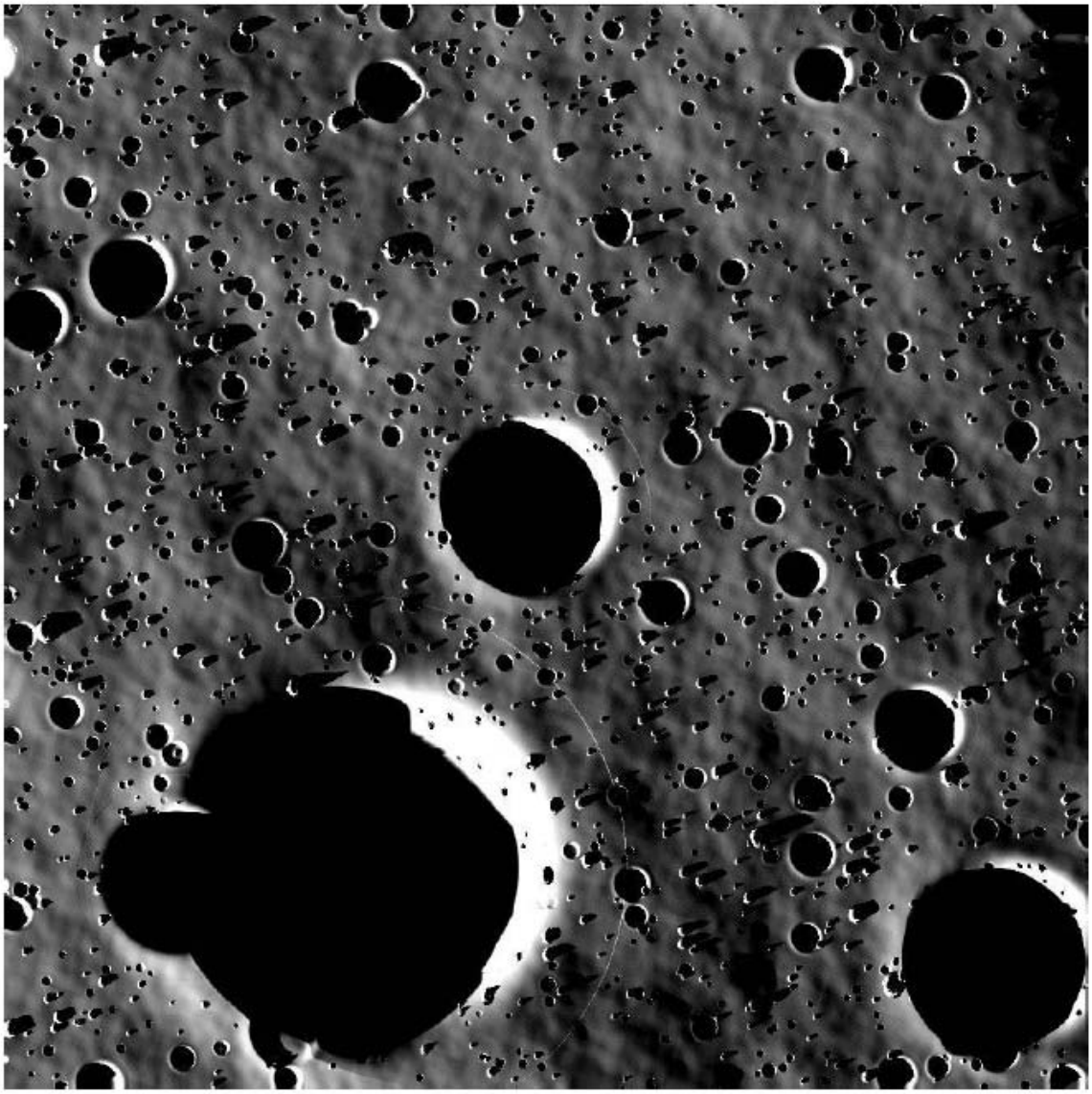


Figure 46: **Expectations photograph for shallow, LIDAR warning.**

Date: _____ Subject #: _____

APOLLO-LIKE: Expectations (AP_S_LI – Vegas)

Please highlight the areas you considered "good" (green highlighter) and "bad" (pink highlighter)



Date: _____ Subject #: _____

APOLLO-LIKE: Display Usage - PFD

Please highlight which aspects of the display you used during LPD. If none, please explain why.

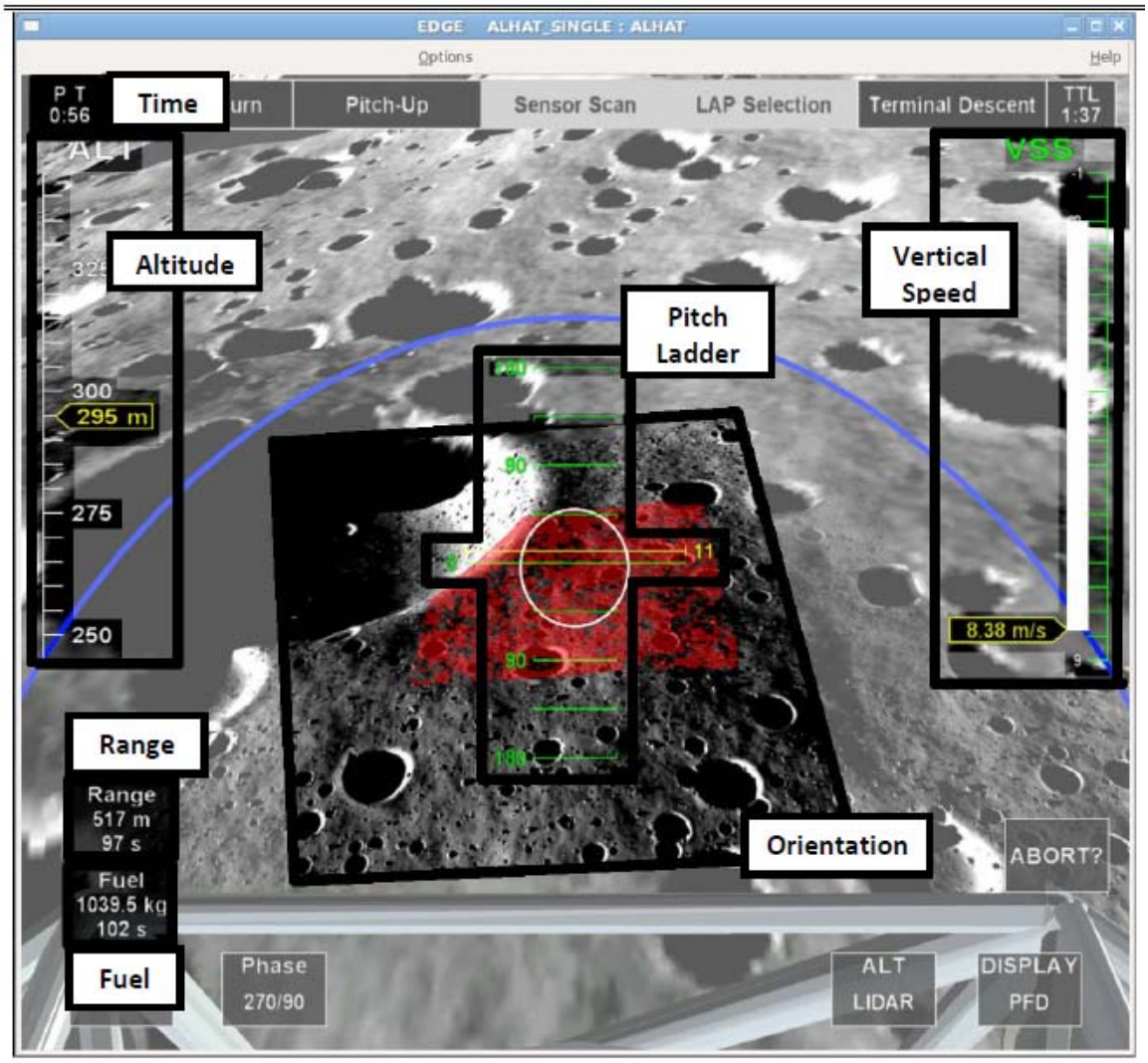
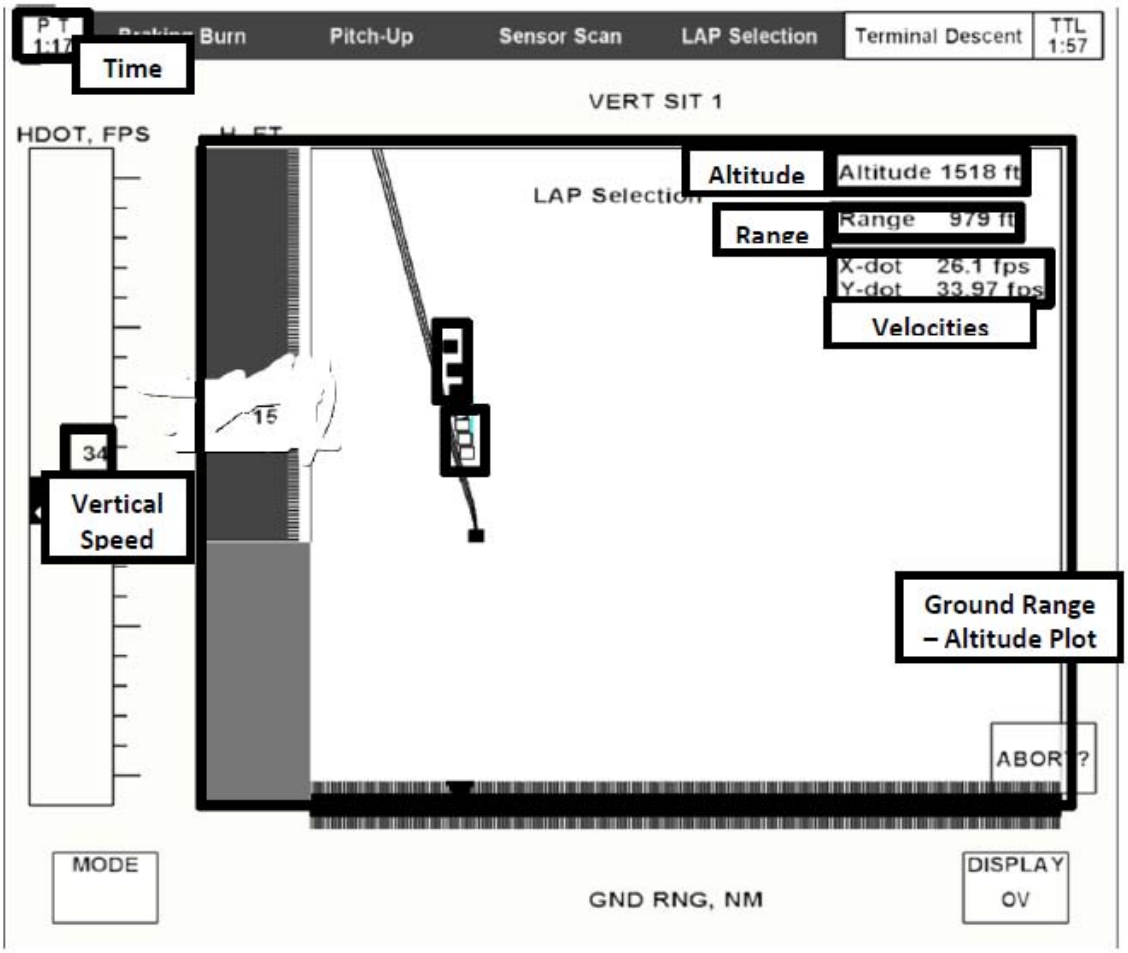


Figure 47: **Primary Flight Display.**

Date: _____ Subject #: _____

APOLLO-LIKE: Display Usage - OV

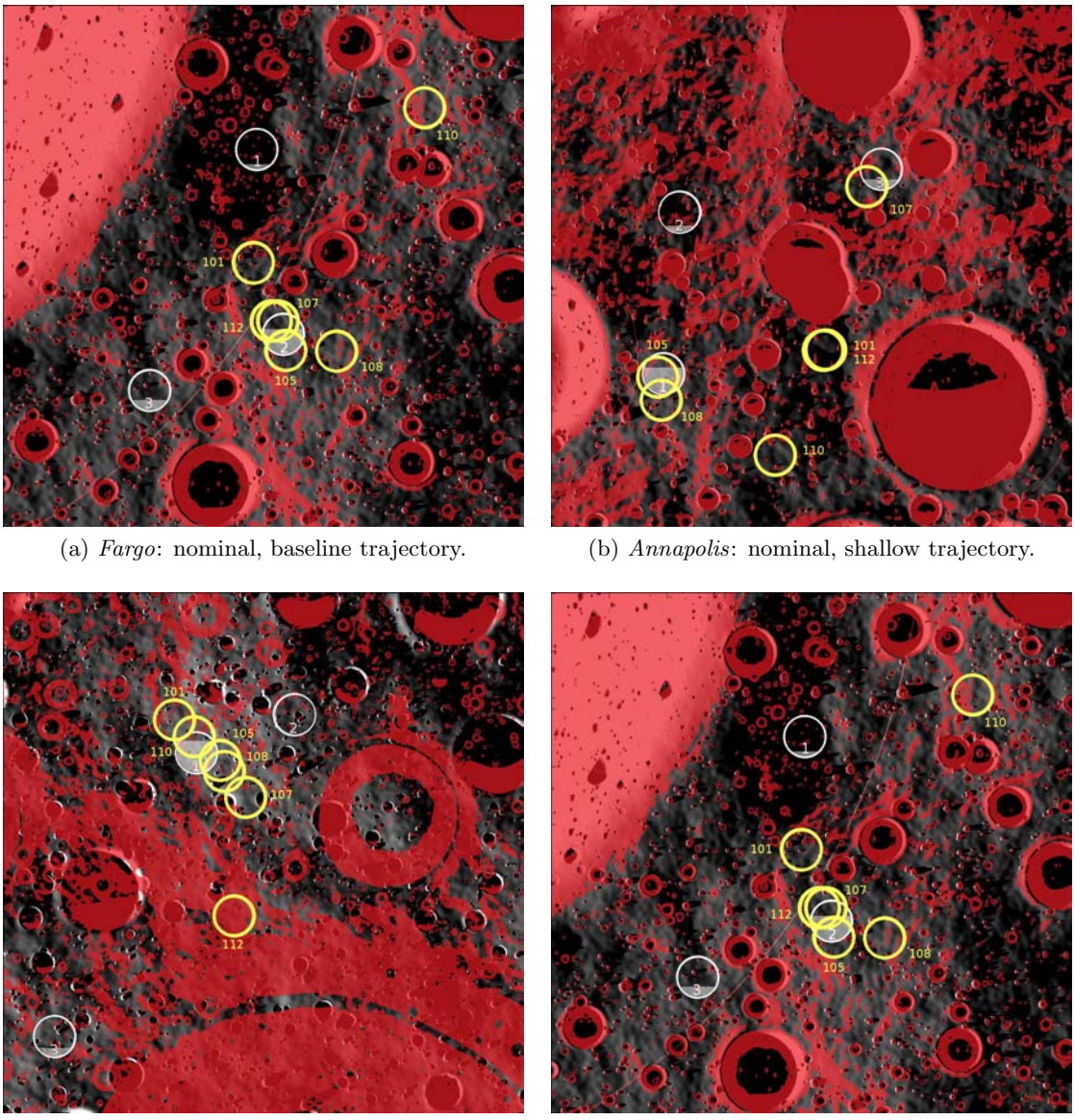
Please highlight which aspects of the display you used during LPD. If none, please explain why.



Did you use the window? If so, what aspects and how?

Figure 48: **Overview display.**

A.2 Distribution of chosen landing sites



(a) *Fargo*: nominal, baseline trajectory.

(b) *Annapolis*: nominal, shallow trajectory.

(c) *Tulsa*: LIDAR warning, baseline trajectory.

(d) *Vegas*: LIDAR warning, shallow trajectory.

Figure 49: **Distribution of chosen landing sites.** The highlighted areas are hazards as identified in the LIDAR scan. The three white landing sites ranked by the moderate automation are shaded according to the percentage of participants who selected this site. The yellow landing sites are those selected by participants with the Apollo-like automation.

A.3 Relative importance of cues per scenario, per participant

This chart illustrates the cues used during each experiment run per participant. For participants using the areal search method, an A is used to represent cues used for the regional, or initial search. B is used to represent cues at the local, or second round of searching. All other cues are marked with an A, for a one-round process. The alphabetical letter is followed by a series of numbers from 1-N, from most important to least important, where N is the number of cues used for that particular scenario.

Participant	Function Allocation	Trajectory Type	Scenario	Strategy Used	Away from Craters	Low Hazard Density	No Hazards in Site	Uniform Color	Not in Darkness	Near POI	In upper 2/3rd	Fuel	Expectations	Work faster	Automation Rank
101	ap	b	li	Areal - Region	A2	A2	A2	A1				A3			
101	ap	b	li	Areal - Site	B1	B1	B1					B2			
105	ap	b	li	Local					A2	A3			A1		
107	ap	b	li	Local			A		A						
108	ap	b	li	Areal - Region		A									
108	ap	b	li	Areal - Site		B1		B2							
110	ap	b	li	Areal - Region		A2		A1			A1				
110	ap	b	li	Areal - Site	B1	B4	B3	B2					B1		
112	ap	b	li	Local			A1	A1					A2		
101	ap	b	no	Areal - Region		A1			A2	A3		A2			
101	ap	b	no	Areal - Site	B1	B1						B2			
105	ap	b	no	Local			A1			A3			A2		
107	ap	b	no	Local			A		A				A		
108	ap	b	no	Areal - Region		A									
108	ap	b	no	Areal - Site		B									
110	ap	b	no	Areal - Region		A2		A1			A1				
110	ap	b	no	Areal - Site	B1	B4	B3	B2					B1		
112	ap	b	no	Local			A1	A1					A2		
101	ap	s	li	Areal - Region	A2	A2	A2	A1	A1			A3			
101	ap	s	li	Areal - Site	B2	B2	B2	B1							
105	ap	s	li	Local					A2	A3			A1		
107	ap	s	li	Local			A		A						
108	ap	s	li	Areal - Region		A									
108	ap	s	li	Areal - Site		B1		B2							
110	ap	s	li	Areal - Region		A2		A1			A1				
110	ap	s	li	Areal - Site	B1	B4	B3	B2					B1		
112	ap	s	li	Local			A1	A1					A2		
101	ap	s	no	Areal - Region	A1	A1	A1			A2		A2			
101	ap	s	no	Areal - Site	B1	B1	B1			B2		B2			
105	ap	s	no	Local			A1			A3			A2		
107	ap	s	no	Local			A		A				A		
108	ap	s	no	Areal - Region		A									
108	ap	s	no	Areal - Site		B									
110	ap	s	no	Areal - Region		A2		A1			A1				
110	ap	s	no	Areal - Site	B1	B4	B3	B2					B1		
112	ap	s	no	Local			A1	A1					A2		
102	mo	b	li	Reranking	A2		A1		A3	A4					

Participant	Function Allocation	Trajectory Type	Scenario	Strategy Used	Away from Craters	Low Hazard Density	No Hazards in Site	Uniform Color	Not in Darkness	Near POI	In upper 2/3rd	Fuel	Expectations	Work faster	Automation Rank
104	mo	b	li	Eliminating	A1	A1		A2							
104	mo	b	li	Eliminating	B1	B1		B2							
106	mo	b	li	Reranking	A2		A2	A2		A1					
109	mo	b	li	Eliminating	A		A	A	A						
111	mo	b	li	Reranking	A1			A2							
113	mo	b	li	Reranking	A1	A1	A1	A1	A2						A2
115	mo	b	li	Eliminating	A2		A1	A1					A2		
102	mo	s	li	Reranking	A2		A1		A3	A4					
104	mo	s	li	Eliminating	A1	A1		A2							
104	mo	s	li	Eliminating	B1	B1		B2							
106	mo	s	li	Reranking	A2		A2	A2		A1					
109	mo	s	li	Eliminating	A		A	A							
111	mo	s	li	Reranking	A1		A3	A2							
113	mo	s	li	Reranking	A1	A1	A1	A1	A2						A2
115	mo	s	li	Eliminating	A2		A1	A1					A2		
102	mo	b	no	Reranking			A1			A2					
104	mo	b	no	Eliminating	A1		A1								
104	mo	b	no	Eliminating	B2		B1								
106	mo	b	no	Reranking	A2		A2	A2		A1					
109	mo	b	no	Eliminating			A		A						
111	mo	b	no	Reranking	A1		A3	A2							
113	mo	b	no	Reranking	A1	A1	A1	A1							A2
115	mo	b	no	Eliminating	A2		A1	A1					A2		
102	mo	s	no	Reranking		A3	A1			A2					
104	mo	s	no	Eliminating	A1		A1								
104	mo	s	no	Eliminating	B2		B1								
106	mo	s	no	Reranking	A2		A2	A2		A1					
109	mo	s	no	Eliminating			A		A						
111	mo	s	no	Reranking	A1		A3	A2							
113	mo	s	no	Reranking	A1	A1	A1	A1							A2
115	mo	s	no	Eliminating	A2		A1	A1					A2		

APPENDIX B

SUPPLEMENTARY MODELING RESULTS

This appendix includes supplementary material related to the validation of the cognitive model.

B.1 Moderate function allocation

Figures 50 and 51 are related to the distribution of cue type and relative importance assignments. Figure 52 illustrates a comparison of cognitive model derived timing estimations and experiment records.

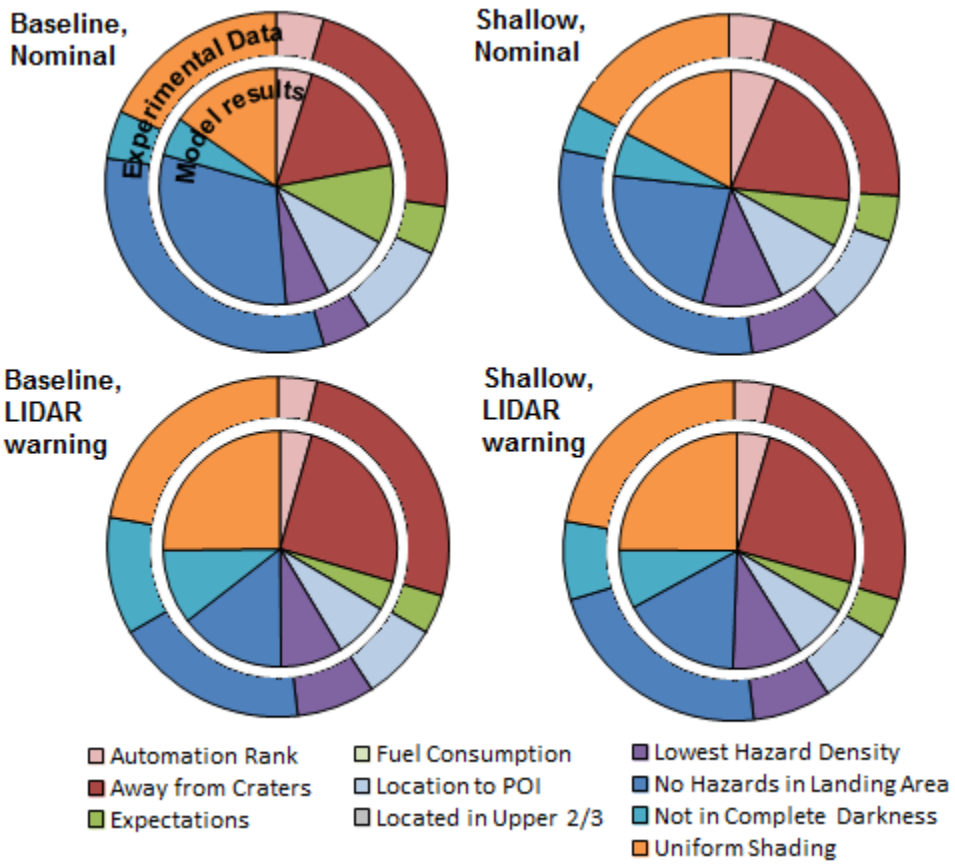


Figure 50: **Comparison of observed frequency of cue use and model selection.** Each of the four quadrants represents information regarding the cue proportionality within the cognitive model and the experiment results. The exterior, larger pie chart is based on experiment results. The interior, smaller pie chart is based off of 1000 cognitive model runs. The close alignment of the two pie charts indicates that the cognitive model resembles experiment results.

B.2 Apollo-like function allocation

Figures 50 and 51 are related to the distribution of cue type and relative importance assignments. Figure 52 illustrates a comparison of cognitive model derived timing estimations and experiment records.

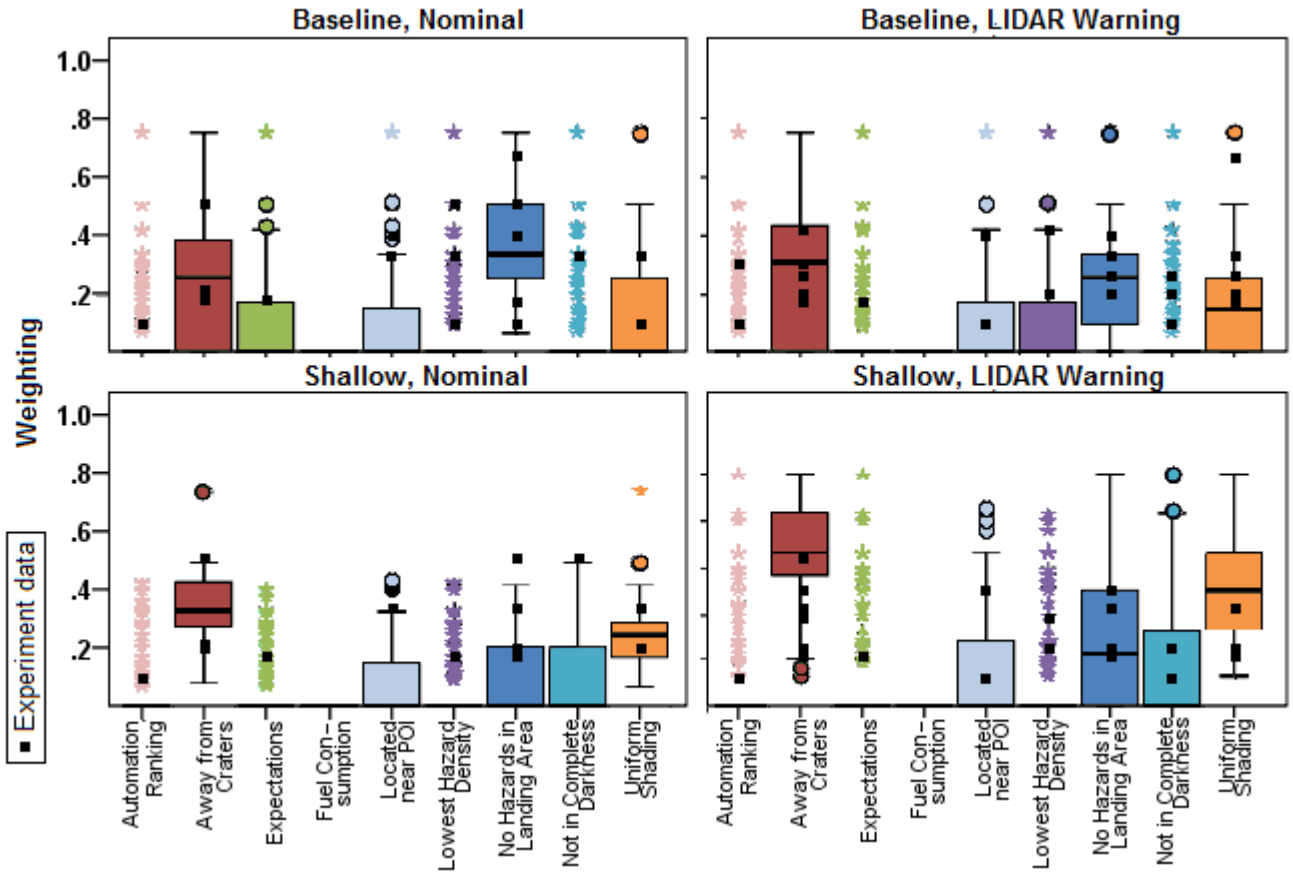


Figure 51: **Comparison of observed frequency of cue weighting and model selection.** The boxplots represent the spread of relative importance (converted to weightings on a 0 - 1 scale) assigned by the cognitive model. The black squares are participant-reported relative importance of the cues. The data is from 1000 runs of the cognitive model.

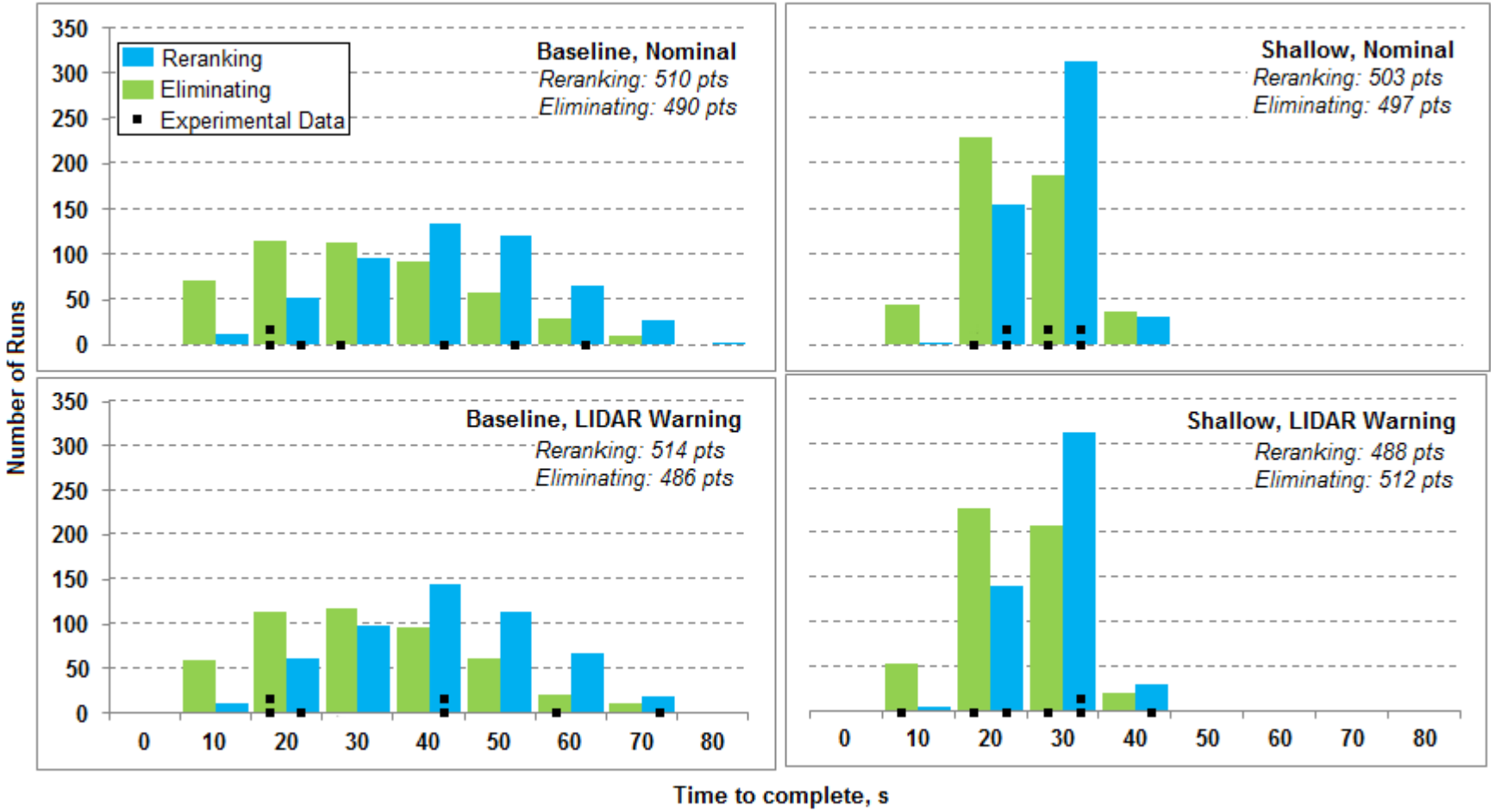


Figure 52: **Comparison of observed frequency of time to complete and model prediction.** The histograms represent the spread of timing estimations as collected in 1000 cognitive runs, grouped by search method. The black squares represent one experiment data point, with multiple points belonging to the same bin stacked on top of each other.

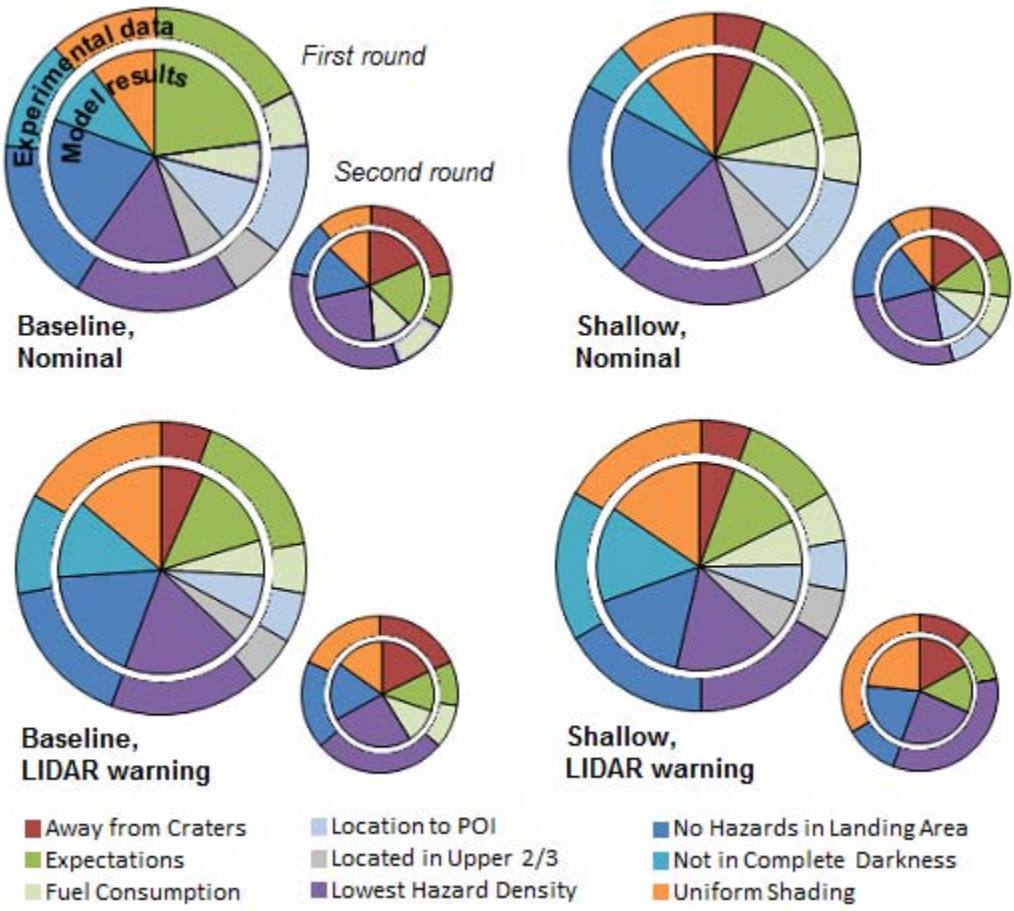


Figure 53: **Comparison of observed frequency of cue use and model selection.** Each of the four quadrants represents information regarding the cue proportionality within the cognitive model and the experiment results. The exterior, larger pie chart is based on experiment results. The interior, smaller pie chart is based off of 1000 cognitive model runs. The close alignment of the two pie charts indicates that the cognitive model resembles experiment results.

Figures 56, 57, 58, and 59 are plots comparing the cognitive model results to actual experiment data. The dark blue region (generally containing a black circle) are top-ranked sites chosen by the cognitive model that are within one landing site of the participant's selected site. The blue region are top-ranked sites chosen by the cognitive model that are *not* within one landing site of the participant's selected site. Green sites are secondary, or second-ranked, sites from the cognitive model. The greyscale backdrop is the cost map of

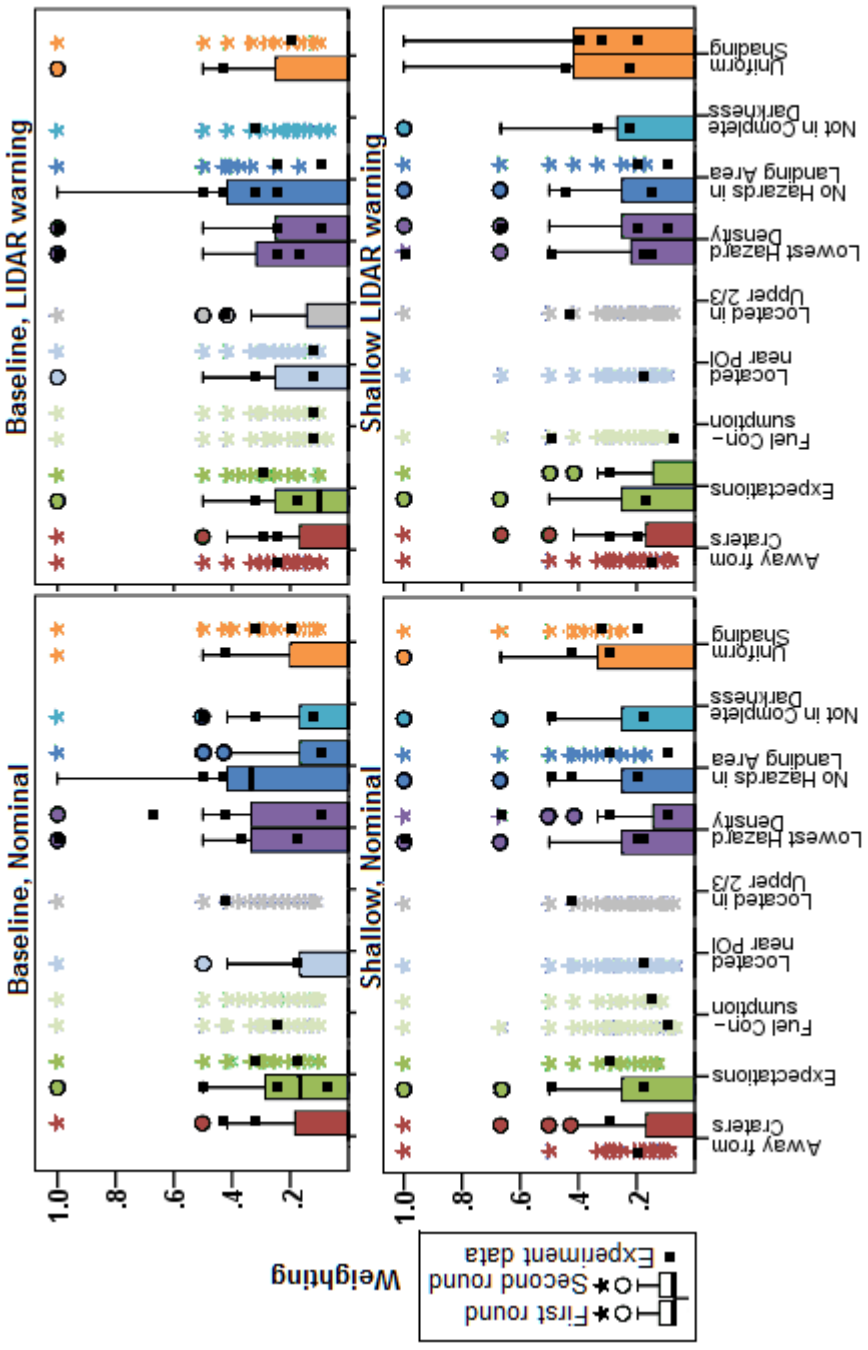


Figure 54: **Comparison of observed frequency of cue weighting and model selection.**

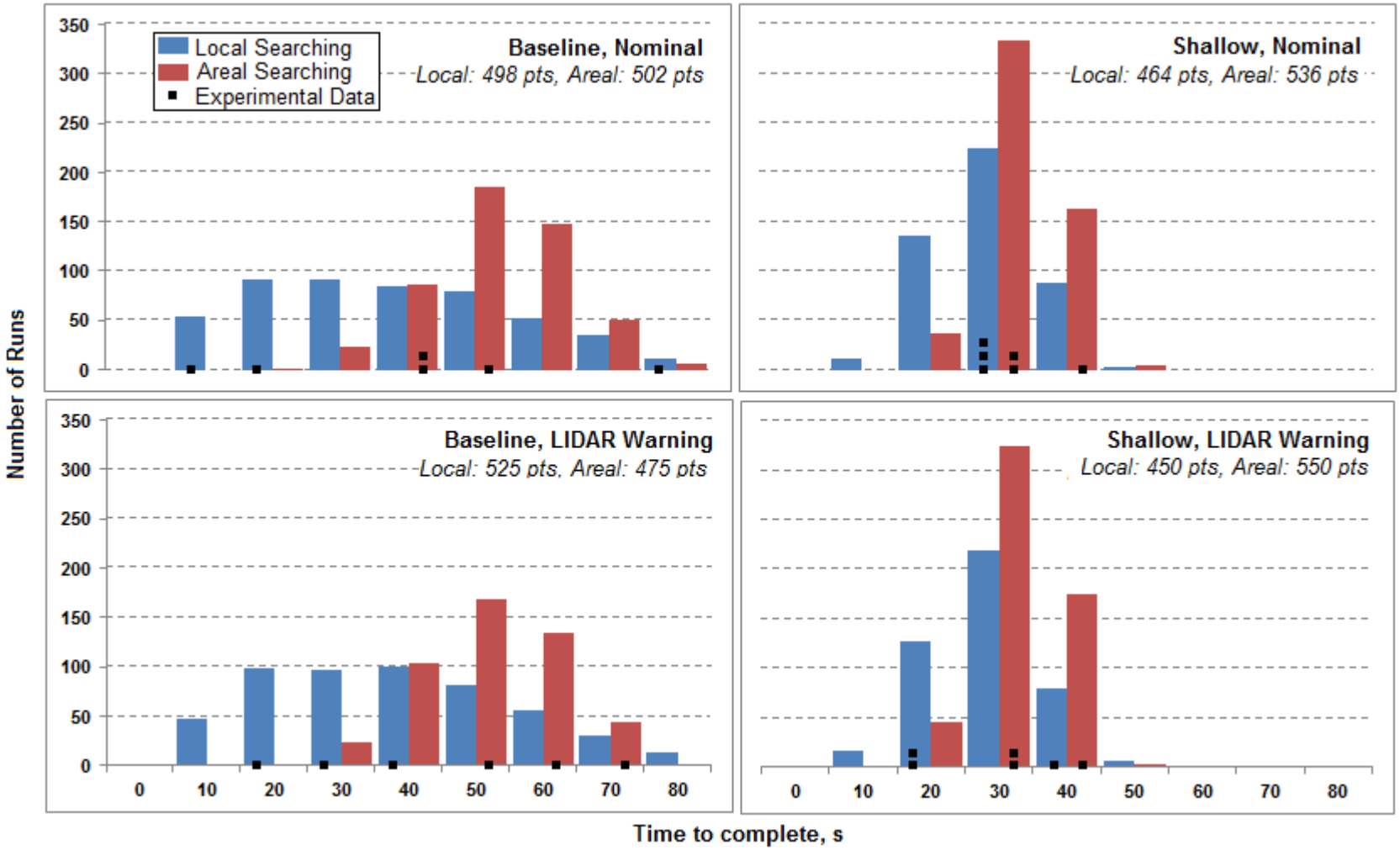


Figure 55: **Comparison of observed frequency of time to complete and model prediction.** The histograms represent the spread of timing estimations as collected in 1000 cognitive runs, grouped by search method. The black squares represent one experiment data point, with multiple points belonging to the same bin stacked on top of each other.

the weighted local and areal, site search criteria. Orange outlines indicate runs of missed detection (cognitive model's secondary site matches participant's selected site); red outlines are for false alarms (neither the cognitive models's chosen or secondary sites match the participant's selected site).

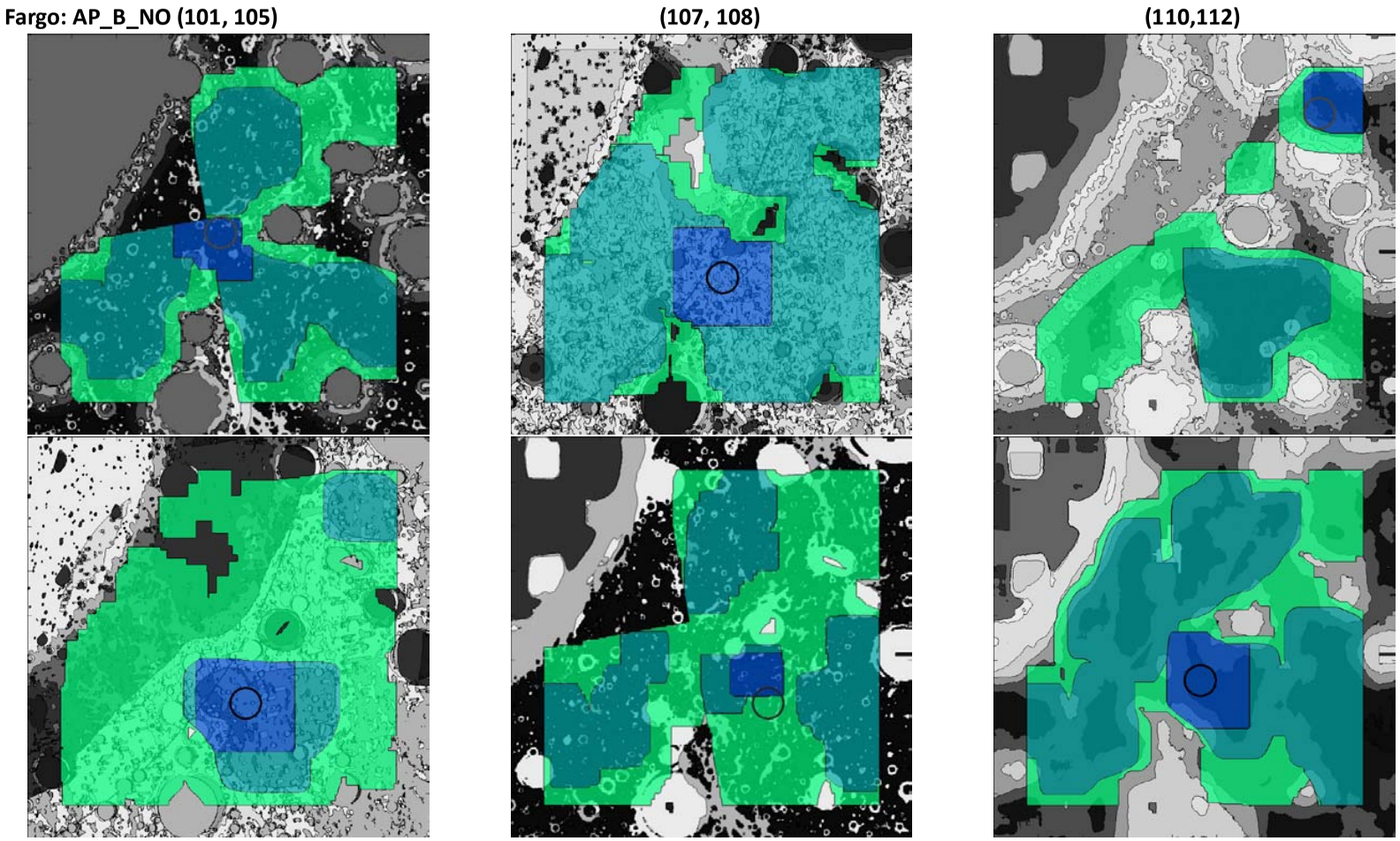


Figure 56: **Individual run validation plots for the Apollo-like function allocation, nominal scenario, baseline trajectory.**

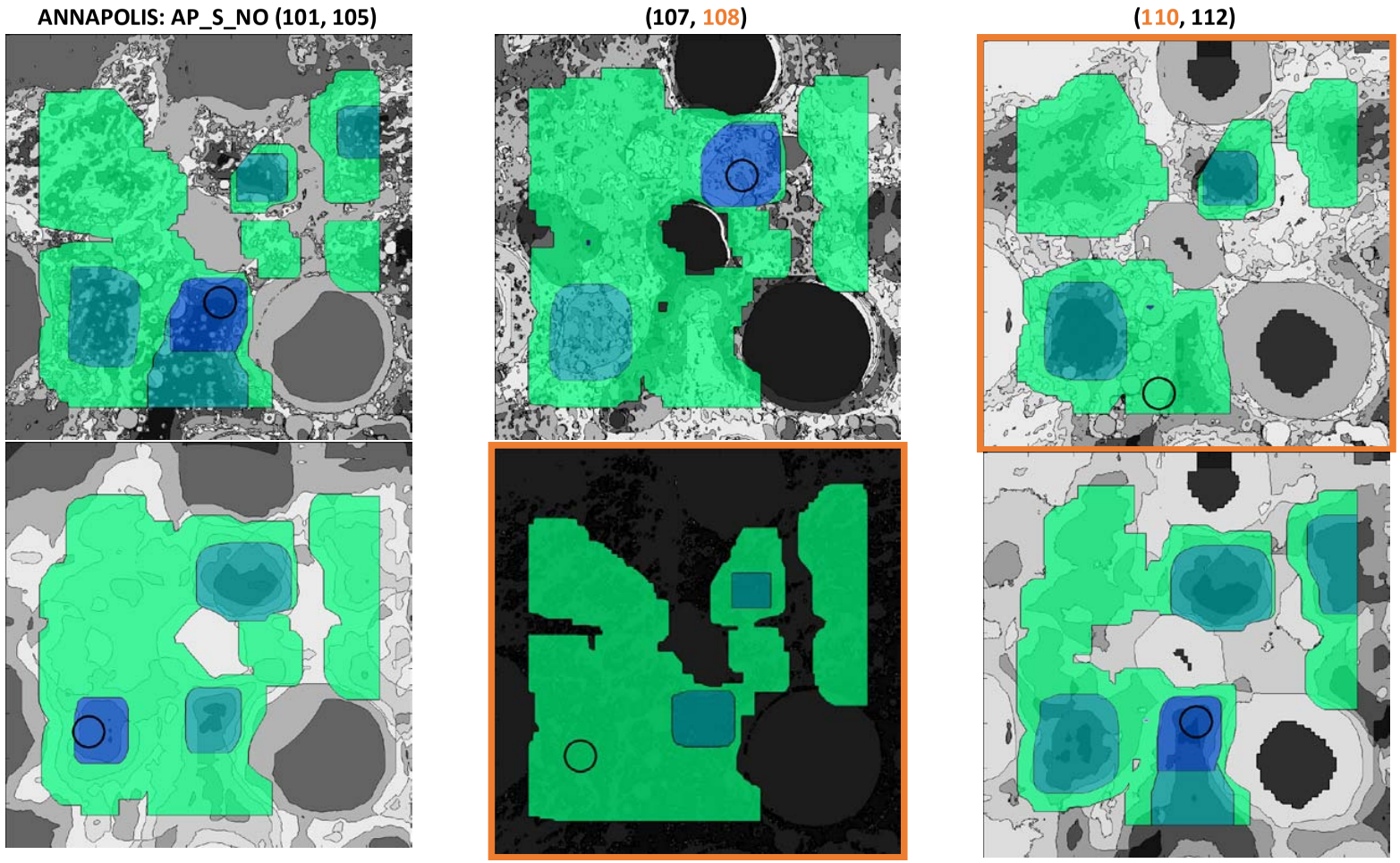


Figure 57: **Individual run validation plots for the Apollo-like function allocation, nominal scenario, shallow trajectory.**

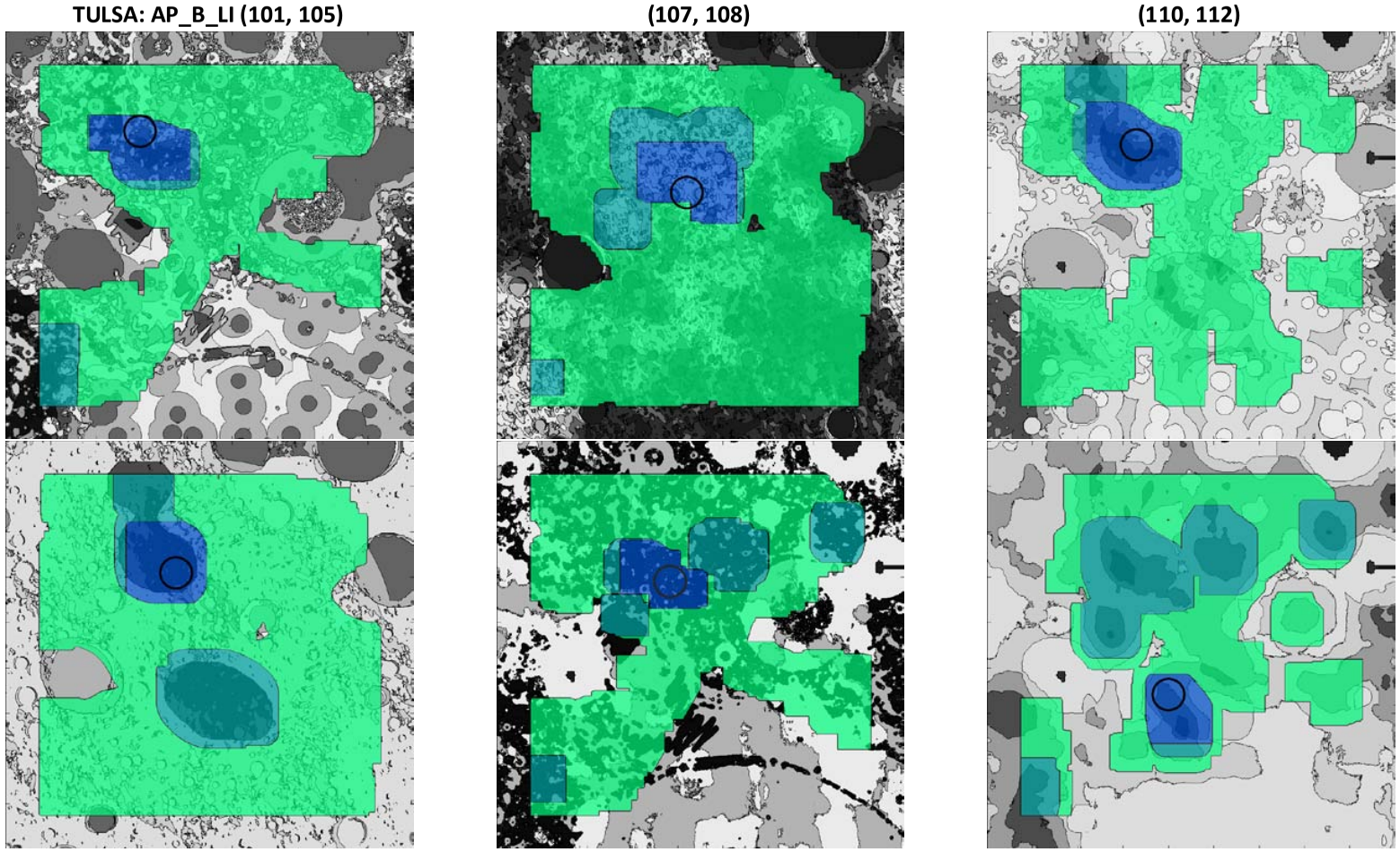


Figure 58: **Individual run validation plots for the Apollo-like function allocation, LIDAR warning scenario, baseline trajectory.**

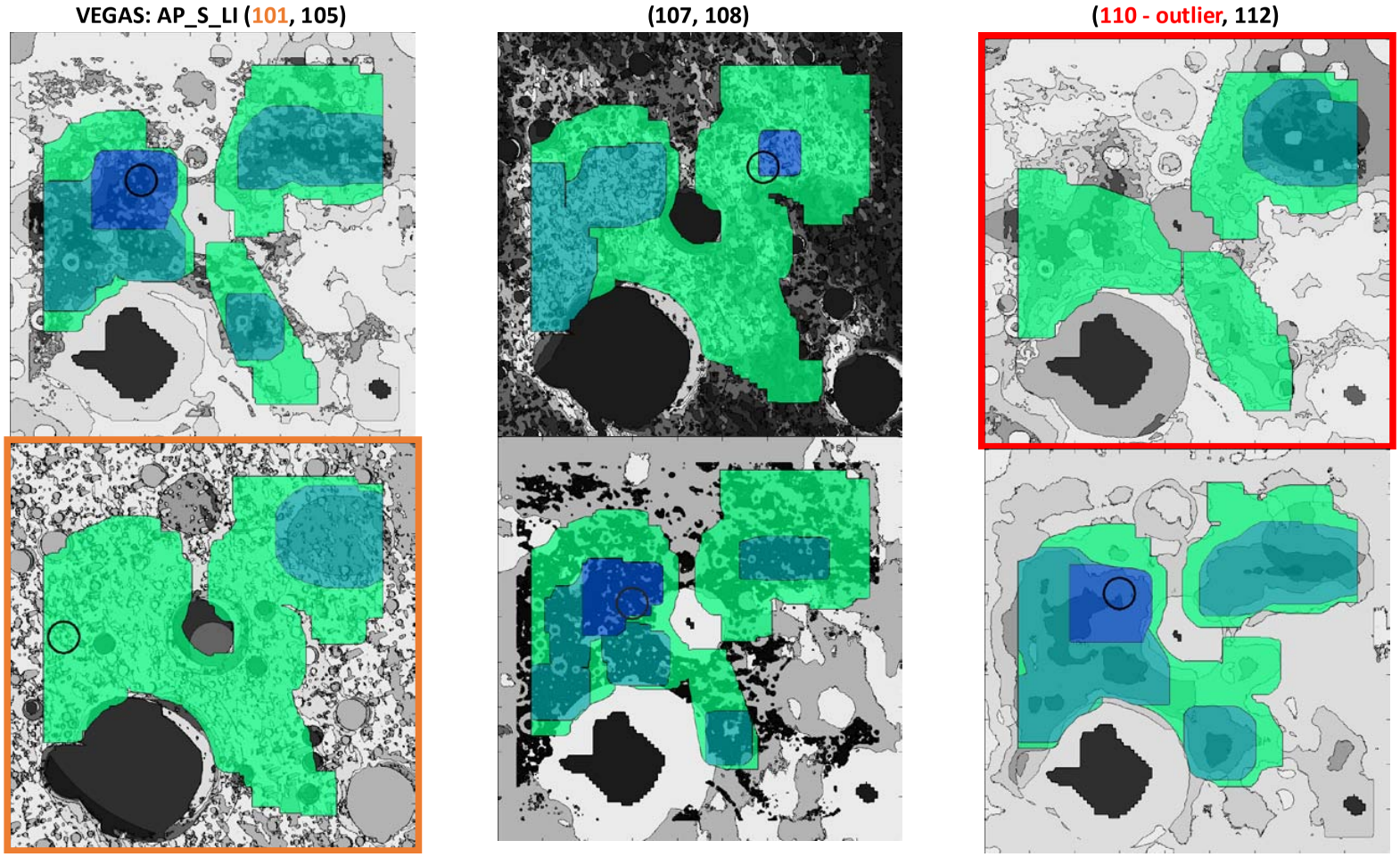


Figure 59: **Individual run validation plots for the Apollo-like function allocation, LIDAR warning scenario, shallow trajectory.**

APPENDIX C

PUBLICATIONS

C.1 Conference Papers

C.1.1 Published Conference Papers

Relevant conference papers:

- **Chua, Z.**; Major, L.; and Feigh, K.; *Modeling Cockpit Interface Usage during Lunar Landing Redesignation*, 15th International Symposium on Aviation Psychology, Dayton, OH, April 2009.
- **Chua, Z.**; and Major, L.; *Task Modeling for Lunar Landing Redesignation*, AIAA InfoTech at Aerospace 2009, Seattle, WA, April 2009.
- **Chua, Z.**; Feigh, K.; and Braun, R.; *Examination of Human Performance During Lunar Landing*, 2010 IEEE Aerospace Conference, Big Sky, MT, March 2010.
- Hirsh, R.; **Chua, Z.**; Heino, T.; Strahan, A.; Major, L.; Duda, K.; *Developing a Prototype ALHAT Human System Interface for Landing*, 2011 IEEE Aerospace Conference, Big Sky, MT, March 2011
- **Chua, Z.**; Steinfeldt, B.; Kelly, J.; and Clark, I.; *System Level Impact of Landing Point Redesignation for High-Mass Mars Missions*, Space 2011, Long Beach, CA, September 2011.
- Feigh, K.; **Chua, Z.**; *Current State of Human Factors in Systems Design*, Human Factors and Ergonomics Society 2011 Annual Meeting, Las Vegas, NV. September 2011. Panel Paper. Panelists: Aragones, A.; Garg, C.; Jacobsen, A.; Rogers, W.; and Shutko, J.
- **Chua, Z.**; Feigh, K. *Integrating Human Factors Principles into Systems Engineering.* 30th Digital Avionics Systems Conference, Seattle, WA, October 2011.

C.1.2 Planned Conference Papers

- **The Effects of Terrain Contrast and Trajectory Design on Low and Moderate Levels of Human Control**. Collaboration with Karen Feigh. Presents the results of the preliminary experiment performed with student pilots. Discusses the effects of terrain contrast, trajectory design, and the Apollo-like and Moderate function allocations on pilot performance. Planned for the International Symposium of Aviation Psychology 2014.

- **Characterizing the Cognitive Process of Astronauts during Landing Point Designation**. Collaboration with Karen Feigh. Presents the results of the main experiment performed with the Astronaut Office. Discusses the effects of terrain contrast, trajectory design, and the Apollo-like and Moderate function allocations on pilot performance. Planned for the Human Factors Annual Meeting 2014.

- **A Computational, Rule-Based Model of Astronaut Decision Making.** Collaboration with Karen Feigh. Presents the development of the cognitive model that was developed based on the results of the main experiment performed with the Astronaut Office. Planned for the Human Factors Annual Meeting 2014.

C.2 Journal Articles

C.2.1 Published Journal Articles

- **Chua, Z.**; Feigh, K.; *Quantifying Pilot Performance during Landing Point Redesignation for System Design*, AIAA Journal of Aerospace Computing, Information, and Communication.

- **Chua, Z.**; Feigh, K.; *Pilot Decision Making during Landing Point Designation*, Cognition, Technology & Work. June 2012.

C.2.2 Submitted Journal Articles

- Wilde, M.; **Chua, Z.**; Fleischner, A.; *Effects of Multi-Vantage Point Systems on the Teleoperation of Spacecraft Docking.* IEEE Transactions on Human-Machine Systems. In Review.

C.2.3 Planned Journal Articles

- **Survey of Cognitive Models.** Surveys the range of primitive models, process models, and cognitive architectures in current literature. Planned submission to Journal of Cognitive Engineering and Decision Making in Spring 2014.

- **Developing a Comprehensive Landing Model.** Discusses the development of the models presented in Chapter 5. Planned submission to IEEE Transactions on Human-Systems in Spring 2014.

- **Systems Analysis of Landing Point Designation.** Discusses the Monte Carlo analysis of the Comprehensive Landing Model and the significance of results. Planned submission to AIAA Journal of Spacecraft and Rockets in spring 2014.

REFERENCES

[1] Brady, T. and Epp, C. D., "Challenge of a Crewed Polar Lunar Landing," Tech. rep., NASA Exploration Technology Development Program, 2008.

[2] Braun, R. D. and Manning, R. M., "Mars Exploration Entry, Descent, and Landing Challenges," *AIAA Journal of Spacecraft and Rockets*, Vol. 44, No. 2, March-April 2007, pp. 310–323.

[3] Feigh, K. M., Dorneich, M. C., and Hayes, C. C., "Towards a Characterization of Adaptive Systems: A Framework for Researchers and System Designers," *Human Factors*, Vol. 54, No. 6, December 2012, pp. 1008–1024.

[4] Bailey, R. W., *Human Performance Engineering: Using human factors/ergonomics to achieve computer system usability*, Prentice Hall, 1989.

[5] Pritchett, A., Kim, S., and Feigh, K., "Measuring Human-Automation Function Allocation," *Journal of Cognitive Engineering and Decision Making*, Vol. Online before print, 2013.

[6] Pritchett, A., Kim, S., and Feigh, K. M., "Modeling HumanAutomation Function Allocation," *Journal of Cogni*, Vol. Online before print, 2013.

[7] Chua, Z. K. and Feigh, K. M., "Integrating Human Factors Principles into Systems Engineering," *Digital Avionics Systems Conference*, AIAA/IEEE, Seattle, WA, October 2011.

[8] Dekker, S. W. A. and Woods, D. D., "MABA-MABA or Abracadabra? Progress on HumanAutomation Co-ordination," *Cognition, Technology and Work*, Vol. 4, 2002, pp. 240–244.

[9] Bainbridge, L., "The Change in Concepts Needed to Account for Human Behavior in Complex Dynamic Tasks," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, Vol. 27, No. 3, May 1997, pp. 351–359.

[10] Fitts, P. M., Chapanis, A., Frick, F. C., Garner, W. R., Gebhard, J. W., Grether, W. F., Henneman, R. H., Kappauf, W. E., Newman, E. B., and Jr, A. C. W., "Human Engineering for an Effective Air-Navigation and Traffic-Control System," Tech. rep., Air Navigation Development Board, Washington, D.C., March 1951.

[11] Chapanis, A., *Human factors in systems engineering*, John Wiley & Sons, Inc., New York, NY, USA, 1996.

[12] Parasuraman, R., Sheridan, T. B., and Wickens, C. D., "A Model for Types and Levels of Human Interaction with Automation," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, Vol. 30, No. 3, 2000, pp. 286–298.

[13] Endsley, M. R. and Kaber, D. B., "Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task," *Ergonomics*, Vol. 42, No. 6, 1999, pp. 462–492.

[14] Sheridan, T. B., *Telerobotics, Automation, and Human Supervisory Control*, MIT Press, Cambridge, MA, 1992.

[15] Proud, R., Hart, J., and Mrozinski, R. B., "Methods for Determining the Level of Autonomy to Design into a Human Spaceflight Vehicle: A Function Specific Approach," *Proceedings of the Performance Metrics for Intelligent Systems Workshop*, 2003.

[16] Boyd, J. R., "The Essence of Winning and Losing," Unpublished Lecture Notes.

[17] Pritchett, A. R., Kim, S. Y., Kannan, S. K., and Feigh, K., "Simulated Situated Work," *Cognitive Models in Situation Awareness and Decision Support*, 2011.

[18] Woods, D., "Cognitive Technologies: the design of join human-machine cognitive systems," *AI magazine*, Vol. 6, No. 4, 1985, pp. 86–92.

[19] Parasuraman, R. and Riley, V., "Humans and Automation: Use, Misuse, Disuse, Abuse," *Human Factors*, Vol. 39, 1997, pp. 230–253.

[20] Smith, P. J., McCoy, C. E., and Layton, C., "Brittleness in the Design of Cooperative Problem-Solving Systems: The Effects on User Performance," *IEEE Transactions on Systems, Man and Cybernetics - Part A: Systems and Humans*, Vol. 27, No. 3, May 1997, pp. 360–371.

[21] Kim, S. Y., *Function Allocation Analysis of Human-Automation Interaction: Development and verification of the work model and the model-based metrics*, Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA, June 2011.

[22] Gray, W. D., Young, R. M., and Kirschenbaum, S. S., "Introduction to This Special Issue on Cognitive Architectures and Human-Computer Interaction," *Human-Computer Interaction*, Vol. 12, 1997, pp. 301–309.

[23] Sun, R., "The Challenges of Building Computational Cognitive Architectures," *Studies in Computational Intelligence*, Vol. 63, 2007, pp. 37–60.

[24] Card, S. K., Moran, T. P., and Newell, A., "The Model Human Processor: An Engineering Model of Human Performance," *Handbook of Perception and Human Performance: Cognitive Processes and Performance*, edited by K. R. Boff, L. Kaufman, and J. P. Thomas, Vol. 2, Wiley, New York, 1986, pp. 1–35.

[25] Miller, G. A., "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *The Psychological Review*, Vol. 63, 1956, pp. 81–97.

[26] Wertheimer, M., *A source book of Gestalt psychology*, Routledge and Kegan Paul, London, 1923.

[27] Gore, B. and Foyle, D., "MIDAS: Man-machine Integration Design and Analysis System," Online, Acessed March 2011.

[28] Gore, B. F., "Man-Machine Integration Design and Analysis System (MIDAS) v5: Augmentations, Motivations, and Directions for Aeronautics Applications," *Human Modeling in Assisted Transportation*, June 2010.

[29] Hammond, K. R., Hursch, C. J., and Todd, F. J., "Analyzing The Components Of Clinical Inference," *Psychological Review*, Vol. 71, 1964, pp. 438–456.

[30] Rothrock, L. and Kirlik, A., "Inferring Rule-Based Strategies in Dynamic Judgment Tasks: Toward a Noncompensatory Formulation of the Lens Model," *IEEE Transactions On Systems, Man, And Cybernetics*, Vol. 33, 2003, pp. 58–72.

[31] Klein, G., *Sources of Power: How People Make Decisions*, MIT Press, MIT, 1998.

[32] Gray, W. D., "Cognitive Architectures: Choreographing the Dance of Mental Operations With the Task Environment," *Human Factors*, Vol. 50, 2008, pp. 497–505.

[33] Johnson, T. R., "Control in ACT-R and Soar," *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, edited by M. Shafto and P. Langley, Lawrence Erlbaum Associates, 1997, pp. 343–348.

[34] Card, S., Moran, T., and Newell, A., *The Psychology of Human-Computer Interaction*, Lawrence Erlbaum Associates, 1983.

[35] Byrne, M. D., *Human Computer Interaction Handbook*, chap. 5, CRC Press, 2007, pp. 94–114.

[36] Lehman, J. F., Laird, J., and Rosenbloom, P., *A Gentle Introduction to Soar, An Architecture for Human Cognition*, University of Michigan, 2260 Hayward, Ann Arbor, MI 48109-2121, 2006, http://ai.eecs.umich.edu/soar/sitemaker/docs/misc/GentleIntroduction-2006.pdf.

[37] "ACT-R: Theory and Architecture of Cognition," Online., 2011, Accesed March 2011.

[38] Campbell, G. E. and Bolton, A. E., *HBR Validation: Integrating Lessons Learned from Multiple Academic Disciplines, Applied Communities, and the AMBR Project*, chap. 10, Lawrence Erlbaum Associates, 2005, pp. 365–396.

[39] Kase, S. E. and Ritter, F. E., "A High-Performance Approach to Model Calibration and Validation," *Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation*, Sundance, UT, April 2009, pp. 39–46.

[40] National Aeronautics and Space Administration, *NASA Space Flight Human System Standard*, vol. 1 ed., September 2009.

[41] National Aeronautics and Space Administration, *NASA Space Flight Human System Standard*, vol. 2 ed., September 2009.

[42] Liskowsky, D. R. and Seitz, W. W., *Human Integration Design Handbook (HIDH)*, Washington, DC, January 2010, Initial Release.

[43] Jr, J. F. P. and West, V. R., *Bioastronautics Data Book: Second Edition. NASA SP-3006*, National Aeronautics and Space Administration, Washington, DC, 2nd ed., 1973, NASA SP-3006.

[44] Department of Defense, *Military Handbook: Human Engineering Guidelines for Management Information Systems*, June 1985, MIL-HDBF-761A.

[45] Department of Defense, *Human Engineering Design Criteria for Military Systems, Equipment, and Facilities*, 1999, MIL-STD-1472F.

[46] Clark, T. K., Stimpson, A. J., Young, L. R., Oman, C. M., and Duda, K. R., "Analysis of Human Spatial Perception during Lunar Landing," *IEEE Aerospace Conference*, 2010.

[47] J O Cappellari, J., "Where on the Moon? An Apollo Systems Engineering Problem," *The Bell System Technical Journal*, Vol. 51, No. 5, May 1972, pp. 955–1127.

[48] Sostaric, R. R., "Powered Descent Trajectory Guidance and Some Considerations for Human Lunar Landing," *AAS Guidance and Control Conference*, No. AAS 07-051, AAS, Breckenridge, CO, February 2007.

[49] Hamza, V., "Lunar Exploration under Earthshine Illumination," Technical Report B69 04017, Bellcomm, Inc., Washington, D.C., April 1969.

[50] Thompson, W. B., "Lunar Landing Site Constraints: The Arguments For and Against One Preselected Site Versus Several Sites," Tech. Rep. TR-64-211-4, Bellcomm, Inc, Washington, DC, January 1964, Also listed under NASA-CR-116911.

[51] Stephen C. Paschall II, Brady, T., Cohanim, B. E., and Sostaric, R., "A Self Contained Method for Safe And Precise Lunar Landing," *IEEE Aerospace Conference 2008*, IEEE, March 2008.

[52] Welford, A., "Attention, strategy and reaction time: A tentative metric," *Attention and Performance IV*, edited by S. Kornblum, Academic Press, New York, 1973, pp. 37–54.

[53] Cavanaugh, J., "Relation between the immediate memory span and the memory search rate," *Psychological Review*, Vol. 79, 1972, pp. 525–530.

[54] Jones, E. M. and Glover, K., "Apollo Lunar Surface Journal," Online, July 2008, accessed 23 September 2008.

[55] Mindell, D. A., *Digital Apollo: Human and Machine in Spaceflight*, The MIT Press, Cambridge, MA, 2008.

[56] Aeronautics, N. and Administration, S., "Astronaut Press Conference," Press Release 60-276, National Aeronautics and Space Administration, Cape Canaveral, Florida, September 1960.

[57] Woods, D., O'Brien, F., Brandt, T., Waugh, L., MacTaggart, K., Vignaux, A., Roberts, I., Wheeler, R., McCray, R., and Jetzer, M., "Apollo Flight Journal," Online, August 2010.

[58] Bennett, F. V., "Apollo Lunar Descent and Ascent Trajectories," Tech. rep., National Aeronautics and Space Administration, Houston, Texas, March 1970, Also presented at AIAA 8th Aerospace Sciences Meeting, January 1970.

[59] Noteborn, R., "Apollo GNC Software," Online., November 2001.

[60] National Aeronautics and Space Administration, Houston, Texas, *Apollo 11 Flight Plan*, July 1969, For AS-506/CSM-107/LM-5.

[61] Major, L. M., Brady, T. M., and II, S. C. P., "Apollo Looking Forward: Crew Task Challenges," *IEEE Aerospace Conference 2008*, No. IEEEAC 1149, IEEE, Big Sky, MT, March 2008.

[62] Cantin, R. and Cantin, J., "Final Approach," Online., Accessed February 2011.

[63] Brady, T. and Epp, C. D., "Operational Concept For the Autonomous Landing and Hazard Avoidance Technology (ALHAT) System," Tech. rep., NASA Exploration Technology Development Program, 2009, ALHAT-2.0-002.

[64] Brady, T., II, S. P., and II, T. P. C., "GN&C Development for Future Lunar Landing Missions," *AIAA Guidance, Navigation, and Control Conference*, No. AIAA 2010-8444, AIAA, Toronto, Ontario, Canada, August 2010.

[65] Davis, J., Striepe, S., Maddock, R., Hines, G., II, S. P., Cohanim, B., Fill, T., Johnson, M., Bishop, R., DeMars, K., Sostaric, R., and Johnson, A., "Advances in POST2 End-to-End Descent and Landing Simulation for the ALHAT Project," *AIAA/AAS Astrodynamics Specialist Conference & Exhibit*, No. AIAA 2008-6938, 2008.

[66] Forest, L., Kessler, L., and Homer, M., "Design of a Human-Interactive Autonomous Flight Manager (AFM) for Crewed Lunar Landing," *AIAA Infotech at Aerospace*, 2007.

[67] Forest, L., Cohanim, B., and Brady, T., "Human Interactive Landing Point Redesignation for Lunar Landing," *Proceedings of the IEEE Aerospace Conference*, Big Sky, MN, March 2008.

[68] Needham, J., *Human-Automation Interaction for Lunar Landing Aimpoint Redesignation*, Master's thesis, Massachusetts Institute of Technology, 2008.

[69] Wen, H. Y., Duda, K. R., Slesnick, C. L., and Oman, C. M., "Modeling Human-Automation Task Allocations in Lunar Landing," *IEEE Aerospace Conference*, 2011.

[70] Wen, H. Y., Johnson, A., Duda, K. R., Oman, C. M., and Natapoff, A., "Decision-Making and Risk-Taking Behavior in Lunar Landing," *PROCEEDINGS of the HUMAN FACTORS and ERGONOMICS SOCIETY 56th ANNUAL MEETING*, 2012.

[71] Chua, Z. K. and Major, L. M., "Task Modeling for Lunar Landing Redesignation," *AIAA InfoTech at Aerospace Conference*, 2009.

[72] Chua, Z. K., Major, L. M., and Feigh, K. M., "Modeling Cockpit Interface Usage During Lunar Landing Redesignation," *International Symposium of Aviation Psychology*, 2009.

[73] Duda, K. R., Johnson, M. C., and Fill, T. J., "Design and Analysis of Lunar Lander Manual Control Modes," *IEEE Aerospace Conference*, 2009.

[74] Major, L. M., Duda, K. R., and Hirsh, R. L., "Crew Office Evaluation of a Precision Lunar Landing System," *IEEE Aerospace Conference*, 2011.

[75] Forest, L., Hirsh, R., and Cohanim, B., "ALHAT Crew Display Survey Guide," *Autonomous Landing and Hazard Avoidance Technology internal presentation*, 2008.

[76] Forest, L., Hirsh, R., and Duda, K., "Results from the Crew Evaluation," Tech. rep., National Aeronautics and Space Administration, December 2008.

[77] Major, L. M., Duda, K. R., Zimpfer, D., and West, J., "An Approach to Addressing Human-Centered Technology Challenges for Future Space Exploration," *AIAA SPACE 2010 Conference & Exposition*, 2010.

[78] Chua, Z. K., Braun, R. D., and Feigh, K. M., "Analysis of Human-System Interaction For Landing Point Redesignation," Masters ae 8900 project, Georgia Institute of Technology, 2009.

[79] Chua, Z. K., Feigh, K. M., and Braun, R. D., "Examination of Human Performance During Lunar Landing," *2010 IEEE Aerospace Conference*, Big Sky, MT, March 2010.

[80] Chua, Z. K. and Feigh, K. M., "Quantifying Pilot Performance during Landing Point Redesignation for System Design," *Journal of Aerospace Computing, Information, and Communication*, Vol. 6, No. 6, 2011, pp. 183–196.

[81] Chua, Z. K. and Feigh, K. M., "Pilot Decision Making during Landing Point Designation," *Cognition, Technology and Work*, Vol. Online, 2012, pp. 1–15.

[82] Roscoe, A. and Ellis, G. A., "A subjective rating scale for assessing pilot workload in flight: A decade of practical use." Tech. Rep. No. RAE-TR-90019, Royal Aerospace Establishment, Farnborough, UK, 1990.

[83] Brunswik, E., "Representative Design And Probabilistic Theory In A Functional Psychology," *Psychological Review*, Vol. 62, 1955, pp. 193–217.

[84] Brunswik, E., "Organismic achievement and environmental probability." *Psychological Review*, Vol. 50, No. 3, 1943, pp. 255 – 272.

[85] Hernandez-Moya, S., "NASA - Innovative Simulation Toolkit for Constructing Simulations," Online, October 2011.

[86] "METECS - 3D Graphics," Online, September 2010.

[87] Hirsh, R. L., Abernathy, M., and Kaiser, V., "Human System Interface Development for ALHAT," *AIAA Space 2009 Conference and Exposition*, Pasadena, CA, September 2009.

[88] Hirsh, R. L., Chua, Z. K., Heino, T. A., Strahan, A., Major, L., and Duda, K., "Developing a Prototype ALHAT Human System Interface for Landing," *IEEE Aerospace Conference*, 2011.

[89] Hirsh, R., "Notes from Evaluation Session," Personal Correspondence.

[90] Chua, Z. K., Steinfeldt, B. A., Kelly, J. R., and Clark, I. G., "System Level Impact of Landing Point Resignation for High-Mass Mars Mission," *AIAA Space 2011 Conference*, 2011.

[91] Johnson, A. E., Huertas, A., Werner, R. A., and Montgomery, J. F., "Analysis of On-Board Hazard Detection and Avoidance for Safe Lunar Landing," *IEEE Aerospace Conference*, 2008.

[92] Johnson, A. E., Keim, J. A., and Ivanov, T., "Analysis of flash lidar field test data for safe lunar landing," *IEEE Aerospace Conference*, 2010.

[93] NASA JSC Public Affairs Office Web Team, "NASA Astronauts," Online., Accessed August 2012.

[94] NASA JSC Public Affairs Office Web Team, "NASA Management Astronauts," Online., Accessed August 2012.

[95] Striepe, S. A., Epp, C. D., , and Robertson, E. A., "Autonomous Precision Landing and Hazard Avoidance Technology (ALHAT) Project Status as of May 2010," *International Planetary Probe Workshop*, 12-18 June 2010.

[96] Cohanim, B. and Collins, B., "Landing Point Designation Algorithm for Lunar Landing," *Journal of Spacecraft and Rockets*, Vol. 46, No. 4, 2009, pp. 858–864.

[97] Fill, T., "Lunar Landing and Ascent Trajectory Guidance Design for the Autonomous Landing and Hazard Avoidance Technology (ALHAT) Program," *Proceedings of the AAS/AIAA Space Flight Mechanics Meeting*, 2010.

[98] Ballard, D. H., "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, Vol. 13, No. 2, 1981, pp. 111–122.

[99] Back, T., *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, 1996.

[100] Tversky, A. and Kahneman, D., "Judgment under uncertainty: Heuristics and biases," *Science*, Vol. 185, 1974, pp. 1124–1131.

[101] Swets, J. A., Green, D. M., Getty, D. J., and Swets, J. B., "Signal detection and identification at successive stages of observation," *Perception & Psychophysics*, Vol. 23, No. 4, 1978, pp. 275–289.

[102] Kramer, L. J., Prinzel, L. J., Bailey, R. E., Arthur, J. J., Shelton, K. J., and Williams, S. P., "Effects of Synthetic and Enhanced Vision Technologies for Lunar Landings," *Digital Avionics Systems Conference*, October 23-29 2009.

[103] Prinzel, L., Kramer, L., Norman, R., Arthur, J., Williams, S., Shelton, K., and Bailey, R., "Synthetic and Enhanced Vision System for Altair Lunar Lander," *International Symposium on Aviation Psychology*, 2009.

[104] Woods, D. D. and Watts, J. C., *Handbook of human-computer interaction*, chap. 26, North Holland, 1997, pp. 616–650.

[105] Johnson, A. E. and Montgomery, J. F., "Overview of Terrain Relative Navigation Approaches for Precise Lunar Landing," *IEEE Aerospace Conference*, March 1-8 2008.

[106] Mourikis, A. I., Trawny, N., Roumeliotis, S. I., Johnson, A. E., Ansar, A., and Matthies, L., "Vision-Aided Inertial Navigation for Spacecraft Entry, Descent, and Landing," *IEEE Transactions on Robotics*, Vol. 25, No. 2, April 2009, pp. 264–280.

[107] GoPro, "GoPro Shows View from Inside Tunnel of Chilean Mine," Electronic, October 2010, Accessed April 2013.

[108] Furihata, H., "Endoscope Camera with Orientation Indicator," February 1975.

[109] Hu, W., Tan, T., Wang, L., and Maybank, S., "A Survey on Visual Surveillance of Object Motion and Behaviors," *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, Vol. 34, No. 3, August 2004, pp. 334–352.

[110] Leger, L. and Bricker, R., "Apollo Experience Report - Window Contamination," Tech. Rep. NASA TN D-6721, Manned Spacecraft Center, Houston, Texas, March 1972.

[111] Pigg, O. E. and Weiss, S. P., "Apollo Experience Report - Spacecraft Structural Windows," Tech. Rep. NASA TN D-7439, Lyndon B. Johnson Space Center, Houston, Texas, September 1973.

[112] Robinson, J. A., Amsbury, D. L., Liddle, D. A., and Evans, C. A., "Astronaut-acquired orbital photographs as digital data for remote sensing: spatial resolution," *International Journal of Remote Sensing*, Vol. 23, No. 20, 2002, pp. 4403–4438.

[113] Riedel, J. E., Vaughan, A. T., Werner, R. A., Wang, T.-C., Nolet, S., Myers, D. M., Mastrodemos, N., Lee, A. Y., Grasso, C., Ely, T. A., and Bayard, D. S., "Optical Navigation Plan and Strategy for the Lunar Lander Altair: OpNav for Lunar and other Crewed and Robotic Exploration Applications," *AIAA Guidance, Navigation, and Control Conference*, No. AIAA 2010-7719, Toronto, Ontario, Canada, August 2010.

[114] Arthur, J. J., Bailey, R. E., Jackson, E. B., Barnes, J. R., Williams, S. P., and Kramer, L. J., "Part-task simulation of synthetic and enhanced vision concepts for lunar landing," *Proceedings of Enhanced and Synthetic Vision*, Orlando, Florida, April 2010.

[115] Morris, A., Goldstein, D., Varghese, P., and Trafton, L., "Plume Impingement on a Dusty Lunar Surface," *AIP Conference Proceedings*, Vol. 1333, 2011, p. 1187.

[116] Manuse, J., "Summary Report ALDAC-2: HRN Methodology for Autonomous Landing and Hazard Avoidance Technology," Tech. Rep. ALHAT-2.0-011, NASA Johnson Space Center, March 2010.

[117] Streefkerk, J. W., Vos, W., and nanja Smets, "Evaluating a Multimodal Interface for Firefighting Rescue Tasks," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56, 2012, pp. 277–281.

[118] Granacki, J., Knapp, D., and Parker, A., "The ADAM Advanced Design Automation System: Overview, Planner and Natural Language Interface," *IEEE 22nd Conference on Design Automation*, 1985, pp. 727–730.

[119] Freitas, R. A., Healy, T. J., and Long, J. E., "Advanced Automation for Space Missions," *7th International Joint Conference of Artificial Intelligence*, Vancouver, British Columbia, Canada, August 1981.

[120] Parasuraman, R. and Miller, C. A., "Trust and etiquette in high-criticality automated systems," *Communications of the ACM*, Vol. 47, No. 4, 2004, pp. 51–55.

[121] Kollar, T., Tellex, S., Roy, D., and Roy, N., "Towards Understanding Natural Language Directions," *5th ACM/IEEE International Conference on Human-Robot Interaction*, 2010.

[122] Knoll, A., Hildebrandt, B., and Zhang, J., "Instructing Cooperating Assembly Robots through Situated Dialogues in Natural Language," *Proceedings of the 1997 IEEE International Conference on Robotics and Automation*, 1997.

[123] Burke, J. D., "Merits of a lunar polar base location," *Lunar Bases and Space Activities of the 21st Century*, Vol. 1, 1985, pp. 77–84.

[124] Angel, R., "A Deep-Field Infrared Observatory Near the Lunar Pole," *Proceedings of the Seventh International Conference on the Exploration and Utilization of the Moon*, 2005.

[125] Bussey, B. J., Spudis, P. D., and Robinson, M. S., "Illumination conditions at the lunar south pole," *Geophysical Research Letters*, Vol. 26, No. 9, May 1999, pp. 1187–1190.

[126] Bussey, D. B., McGovern, J. A., Spudis, P., Neish, C., Noda, H., Ishihara, Y., and Sorensen, S.-A., "Illumination conditions of the south pole of the Moon derived using Kaguya topography," *Icarus*, Vol. 208, 2010, pp. 558–564.

[127] Dale, S. and Cooke, D., "Exploration Strategy and Architecture," *2nd Exploration Conference*, Vol. 6, Houston, TX, 2006.