Quantitative Assessment of Human Control on Landing Trajectory Design

Zarrin K. Chua

December 2, 2010

Contents

Nomenclature

α	Type I error
T_c	Time to Complete
AFM	Autonomous Flight Manager
ALHAT	Autonomous Landing and Hazard Avoidance Technology
DSKY	Display and Keyboard unit
LEM	Lunar Excursion Module
LIDAR	LIght Detection And Radar
LMP	Lunar Module Pilot
LPD	Landing Point Designation
LSS	Landing Site Score
PFD	Primary Flight Display
PPS	Pilot Performance Score

1 Introduction

An increased thirst for scientific knowledge and a desire to advance humanity's presence in space prompts the need for improved technology to send crewed vehicles to places such as the Moon, Mars, and nearby passing asteroids [2]. Landing at any of these locations will require vehicle capabilities greater than that previously used during the Apollo program or those applied in Low Earth Orbit. In particular, the vehicle and the on-board crew must be capable of executing precision landing in sub-optimal landing conditions during time-critical, high-stakes mission scenarios, such as Landing Point Designation (LPD) , or the critical phase of determining the vehicle's final touchdown point. Most proposed solutions involve automated control of landing vehicles, accepting no input from the on-board crew - effectively relegating them to payload [11], [30]. While this method is satisfactory for some missions, an automation-only approach during this critical mission phase may be placing the system at a disadvantage by neglecting the human capability of [what?]. Therefore, the landing system may result in a lack of dynamic flexibility to unexpected landing terrain or in-flight events.

It is likely that executing LPD will require an ideal distribution of authority between the on-board crew and an automated landing system. However, this distribution is application-specific and not easily calculated. Current science does not provide enough detailed or explicit theories regarding allocation of automation, and the advantages provided by biological and digital pilots (either acting as the sole authoritarian or as a coordinated team) are difficult to describe in quantitative measures. Despite previous experience in piloting vehicles on the Moon, few cognitive models describing the decision-making process exist. The specialization of the pilot and the application pose significant practical challenges in regular observations in the target environment.

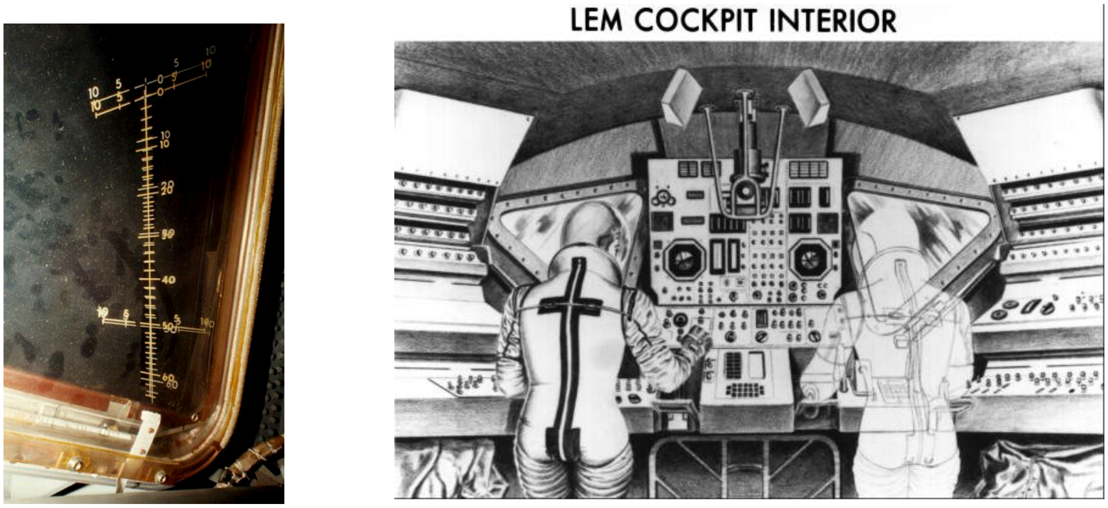
The lack of quantitative knowledge results in predominantly qualitative design trade-offs during pre-mission planning. While qualitative analyses have proven to be useful to the mission designer, an understanding founded on quantitative metrics regarding the relationship between human control and mission design will provide the sufficient supplementary information necessary for overall success. In particular, increased knowledge of the impact of human control on landing trajectory design would allow for more efficient and thorough conceptual mission planning. This knowledge would allow visualization of the flight envelope possible for various degrees of human control and help establish conceptual estimations of critical mission parameters such as fuel consumption or task completion time. This report details an experiment undertaken to further understanding of the impact of moderate degrees of human control on landing trajectory design or vice versa during LPD. This report briefly summarizes current understanding and modeling of moderate control during LPD and similar applications, reviews previous and current efforts in implementing LPD, examines the pilot study to observe subjects in a simulated LPD task, and discusses the significance of findings from the pilot study.

2 Background

The relationship between moderate degrees of human control and landing trajectory design during landing point designation (LPD) can be broken into several elements: the definition of LPD; quantifying the allocation of responsibilities for completing LPD; and cognitive modeling efforts in LPD and related fields. This section describes each of these elements in greater detail and relates to previous and current studies.

2.1 Landing Point Designation

Landing Point Designation (LPD), in its simplest form, is the opportunity to select a final landing site. This decision-making process trades off selection criteria such as the fuel required to reach a site, safeness of the site (*e.g.*, the roughness, slope, and proximity of rocks and craters), or the proximity of the landing to a point



(a) Landing Point Designator (b) LEM Cockpit Interior

Figure 1: Landing Point Designation for Apollo [17].

of interest. The relative importance of each criterion is also determined by the operator and is based on *a priori* or real-time data. This task can be completed in a variety of ways: an on-board pilot, an automatic landing system, a remote operator, or a combination of these entities. There have been six instances of lunar LPD (the Apollo landings) but this task has been compared to other applications such as engine out [6]. Lunar LPD generally occurs during the final approach of the landing trajectory, prior to terminal touchdown and lasts no more than two minutes.

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 Lunar Excursion Module (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) [22] participated in LPD, with very specific roles. The Commander worked primarily with the landing point designator (Figure 1). 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. 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. This process continued for several iterations, with the LMP reading to until the LEM reached the predesignated time-to-go and performed terminal descent.

The crew interacted with the on-board flight computer through the Display and Keyboard unit (DSKY). The crew used 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 [24]. 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.

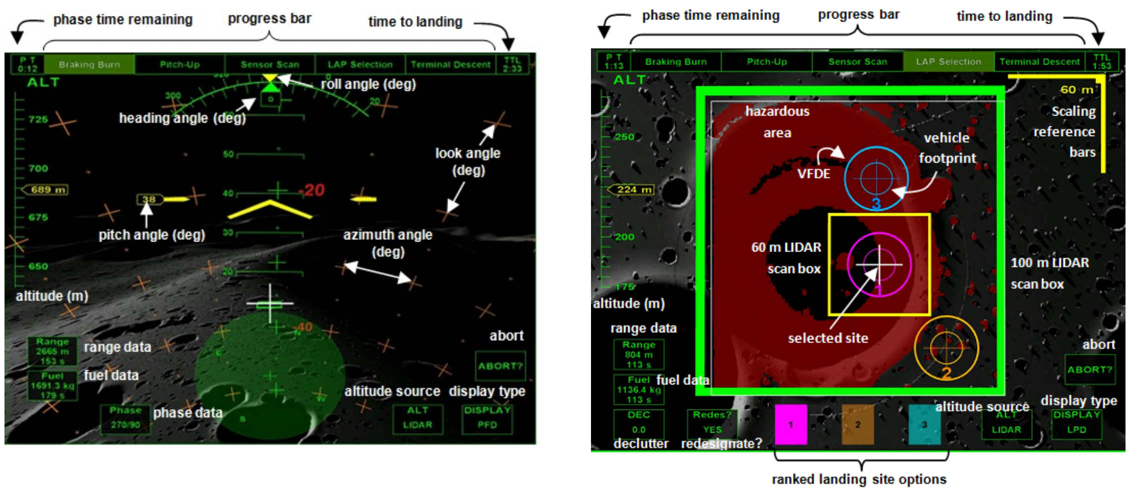
Unsurprisingly, this method of LPD places a strong reliance on the landing trajectory. The LEM must fly at an angle that permits visibility of the landing site. This viewing angle is a function of the landing location and the sun angle. At certain angles, craters and rocks may cast shadows across the terrain that distort the astronaut's view, either by hiding obstacles or obfuscating landmarks that the astronauts used for terrain orientation. The effects of lighting were particularly pronounced in Apollo 12. The astronauts relied on the "Snowman" crater configuration, but temporarily could not find this landmark, adding a slight delay to the mission sequence. In general, the LEM flew at a flight path angle of n° with a viewing angle of n°. LPD began after Powered Descent Initiation and shortly after entering "high gate" (P64) [32]. At high gate, the LEM was at a range of 26000 ft [24], an altitude of 7515 ft, had a descent rate of -145 fps, and an inertial velocity of 506 fps [3]. At the end of LPD, the vehicle entered "low gate" (P65/P66) at a range and altitude of 2000 ft [24] and 500 ft, respectively, with a descent rate of -16 fps, and an inertial velocity of 55 fps [3].

During the Apollo era, standard perceptions of piloting changed substantially. The Apollo astronauts were converts of the test pilot program, accustomed to piloting new aircraft through standard procedures and the occasional death-defying maneuver. A pilot's worth was defined predominantly by his skill at mastering the machine, with little guidance from any automation system. However, the risks and unknowns surrounding the first manned landing attempt on another surface prompted the development of a robust digital autopilot or flight computer, leading several pilots to feel as though they were "spam in a can" [21]. 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 [1]." This feeling exists today, and has survived through the modern era of spaceflight. However, as the situations have grown increasing complex, this perception has adapted. Modern approaches to landing on the Moon has increased the role of the automation system. This change is most apparent in the Autonomous Landing and Hazard Avoidance Technology (ALHAT) project.

The ALHAT project is sponsored by NASA and involves developing a technology suite to provide safe and precise lunar landing for both crewed and robotic vehicles [4]. This technology aims to provide global access to the lunar surface under any lighting condition. An extensive study has been undertaken to design the appropriate displays to support astronauts during LPD. In this investigation, LPD is conducted similarly to the Apollo mission: the task occurs after Powered Descent Initiation, during the final approach to the lunar surface. The vehicle is at an altitude of 1000m with a velocity of 100 m/s at the start of LPD, and terminal descent occurs at about 30m from the surface [11]. However, there are variations on this trajectory, depending on the mission.

There are several notable differences in the LPD methodology. First, ALHAT employs a LIght Detection And Radar (LIDAR) sensor to capture real-time data regarding the terrain. Second, an Autonomous Flight Manager (AFM) processes this data and selects alternative sites to present to the operator. Third, the crew does not have the option of manually piloting the vehicle to the lunar surface. Although windows are expected to be on the final vehicle design, the majority of the vehicle interaction is through the AFM and the flight displays. There are three displays available to the on-board crew, but only the primary flight display (PFD) and the LPD display (Figure 2) provide information to the crew to complete LPD. The PFD provides vehicle status information - fuel remaining, attitude, rate of descent, altitude, etc - but also provides a camera viewpoint of the lunar terrain in the foreground. The LPD display presents a shaded top-down picture of the landing area, with the results of the LIDAR scan superimposed. Areas that are hazardous (craters, rocks, slopes, or roughness values exceeding tolerances) are shaded red, and three ranked alternative sites are provided. The crew selects a new landing site by pressing the associated button, and the AFM sends the appropriate guidance commands to the vehicle.

At this time, the AFM automatically pilots the vehicle to the ground. However, if the crew does not make a decision within the allotted time, the AFM will default to the first ranked site.



(a) Primary Flight Display.

(b) LPD Display.

Figure 2: Reference Displays Used for ALHAT.

A full list of symbology can be found in the Human-System Interface Cockpit Display Documentation [5].

2.2 Allocation of Responsibilities

The LPD task can be completed by any of the following operators: an on-board crew, an automatic landing system, or a remote operator. In each case, the identified operator is in charge of all aspects of LPD with no interference from the other operators. Additionally, the type of LPD can then be easily described - *fully manual*, *robotic*, *tele-operated*, respectively. However, when two or more operators are working in tandem, describing the type of LPD performed is more difficult. In particular, determining the absolute amount of automation allocation is a challenge as only relative descriptors (*i.e.*, more automated, more manual) can be used. While knowing an absolute value may not be critical to mission operations, this lack of a ranking system reduces the ability to draw inferences regarding an increased application of automation. For example, a common concern is the correlation between automation and human workload. Simplistically, the amount of automation could be measured in terms of a percentage of the task that is assigned to the automation. General intuition leads to the belief that a greater percentage of automation requires less workload from the human operator. However, current literature [31] [28] notes that the category of task that is automated must be taken into consideration. Human operators perform better if information collection and presentation is automated; issues such as complacency and loss of situation awareness become a problem if the automation plays a large role in the decision-making process.

LPD is composed of several fundamental subtasks: collecting information, interpreting the information, creating solution options, selecting a solution, executing the chosen solution. The information necessary for LPD consists primarily of the vehicle status, mission status, and characteristics of the landing area terrain. Once this information is collected, it must be interpreted, creating a cost map of desirable and undesirable terrain based on mission constraints and operator preferences. From this cost map, landing site options are created in response to the problem of where to land. A final landing site is selected and executed by piloting the vehicle to the final destination. Each type of LPD, Apollo, and ALHAT can be compared to this schema to examine where the automation or crew is assigned. Tele-operated LPD is omitted in Table 1 as it is unlikely due to delays in communication.

The chart illustrates the complexity of trying to rank the amount of automation. Is each task assumed to be of equal importance or difficulty? Are all combinations of n automated tasks, $5 - n$ human tasks equatable? Several studies [26] [27] [19] [25] have created taxonomies of levels of automation for general

Operator	Collecting	Interpreting	Creating	Selecting	Executing
On-board crew	Human	Human	Human	Human	Human
Apollo	Human/Auto	Human	Human	Human	Human/Auto
ALHAT	Human/Auto	Auto	Auto	Human/Auto	Auto
Automatic	Auto	Auto	Auto	Auto	Auto

Table 1: LPD task breakdown.

cases. Endsley and Kaber's ten-point levels of automation scale provides sufficient principles appropriate to define automation and compare differences. This scale and its application to the four options (on-board crew, Apollo, ALHAT, and Automatic) is listed in Table 2.

Level	Monitor	Generate	Select	Implement	LPD Option
1: Manual Control	H	H	H	H	On-board crew
2: Action Support	H/C	H	H	H/C	Apollo
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	ALHAT
7: Rigid System	H/C	C	H	C	ALHAT
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	Automatic

Table 2: Endsley and Kaber's Ten-Point Levels of Automation.
The allocation of responsibilities is indicated by Human (H) or Computer (C).

Under this taxonomy, *monitoring* is regarded as scanning displays/terrain to perceive system status; *generating* consists of formulating landing site options; *selecting* is choosing a final landing site; *implementing* is piloting the vehicle to the targeted destination. As seen in Table 2, the ALHAT version of LPD does not align perfectly with only one level of automation. The AFM does not permit the on-board crew to generate any site selections, unlike Level 6, but the crew is not the sole authority on site selection, which is the definition of Level 7. As seen in both tables, defining an absolute amount of automation for LPD is impossible, but the allocation of responsibilities can still be regarded as a rank variable in either levels of automation or levels of human control based on the resemblance of one option to an extreme. However, as the changes of allocation become less pronounced, the more difficult it is to rank the options. Nevertheless, a level of automation comparable to Apollo is universally considered as a "high" level of human control whereas an automation allocation such as ALHAT could be considered as a "moderate" level (with respect to both extremities).

2.3 Cognitive Modeling

Much of the current understanding of the relationship between human control and landing trajectory design during LPD is limited to six data points - the performance of the only six lunar landing attempts, all from the Apollo era. Although this number is low by statistical and modeling standards, much can be gleaned from the LPD task description, training material, astronaut debriefings, related scenarios, and simulated studies. There are several studies of relevance to this investigation. Bennett [3] conducted an investigation following Apollo 11 comparing the pre-mission planning and actual landing trajectory of Apollo 11. This study also included commentary on Apollo 12. In these analyses, Bennett noted that pilot control resulted in variations in the flight profile; this was explained by Neil Armstrong as "I [was] just absolutely adamant about my God-given right to be wishy-washy about where I was going to land." Other historical studies involve the Space Shuttle Orbiter. Holland and VanderArk [16] performed a task analysis regarding Shuttle re-entry. In this study, they examined the roles of the three crew members involved with re-entry. Their detailed investigation included vehicle status at each subtask and a list of interactions with the Orbiter interface. However, this analysis lists the tasks in a sequential order and does not account for off-nominal situations.

The ALHAT team has conducted studies regarding the crew during LPD, including prompts for supervisory commands [15], suggestions for cockpit display designs [14], in particular, for the primary flight displays and horizontal situation indicator.

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 [13]. Needham investigated the impact of varying levels of automation on human performance during LPD, concluding that higher automation allowed for quicker time to complete [23]. 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. Sostaric and Rea modeled the impact of LPR on trajectory design [29]. This impact exists primarily in the need for a window viewing angle and required vehicle divert capability. This study provides initial estimations on the change in metrics (e.g., time to touchdown, trajectory profile) but only for a high-level concept of LPD; no further commentary on specific human tasks is provided. Lastly, Chua *et al.* have derived a task model and used this model to examine bottlenecks of LPD [8]. These bottlenecks were addressed by redesigning the LPD display to simplify the information layout and to utilize new symbolism to represent site characteristics. This LPD task model also incorporates expert decision-making theory [18] to account for specialized astronaut behavior [9]. However, this study is based on theory and lacks observations from equatable subjects.

Other studies, such as Chua and Feigh [7] utilized experimental data to quantify human performance during LPR under varying definitions of performance success. The results of this study illustrate that pilots' decision-making times result in heavy penalties with regard to fuel consumption. However, pilots generally select sites that are regarded as safe or near the points of interest. The analysis results were also compared to a reference automation system to determine areas of better performance. The automated landing system generally performed better than pilots in missions where fuel was the most critical factor. However, the speed of the automation did not necessarily translate to better sites, as pilots routinely selected safer sites closer to points of interest.

3 Experimental Methods

3.1 Objective

As part of a larger body of research on characterizing the LPD performance envelope, especially with changes to landing trajectory and lunar lighting effects, this paper presents the results of an experiment designed to measure the impact of lunar lighting on pilot performance during LPD, especially during moderate levels of human control. In addition to this main research question, the results of this experiment may provide answers as to what information the pilots use to inform the decision-making process and what strategies are employed to select the best landing site. Crew performance and decision-making are likely influenced by several factors. Previous lunar landings and consultation with subject matter experts have identified *lunar lighting* as the most paramount influence on crew performance. This experiment establishes lighting as an independent variable with two treatments: *ideal* which represents conditions in which craters are easily outlined and not distorted by shadow or conversely, lit too brightly there is low contrast discernible to the human eye; and *poor* which provides minimal lighting (enough to significantly affect pilot perception) but excludes flying in completely dark conditions (*i.e.*, the entire landing area must be mostly visible, not a black square). Quantitatively, maps with approximately 25% shadowed area were deemed “poor”; greater percentages of map shadows were excluded from consideration in this experiment.

Three main hypotheses have been formulated: as the lighting decreases in quality (ideal to poor), the LPD task grows more difficult and thus 1) increases the task time to complete; 2) decreases the landing site score; and 3) decreases the pilot performance score. Two additional sub-hypotheses were formulated regarding the pilot’s perception of the task: 1) as pilot trust in automation increases, the landing site score and pilot performance score improve; and 2) as pilot perception of lighting is factored more into the decision-making process, the landing site and pilot performance scores decreases due to the inclusion of this additional information.

Five dependent measures were selected to test the formulated hypotheses and a sixth variable was measured to provide additional information regarding the LPD scenario context. (1) *Time to complete* (T_c) is the time in seconds from the beginning of the LPR task, when the pitch-up maneuver is completed, to when the vehicle touches the lunar terrain. Based on the experiment procedure, T_c is measured from when the participant initiates the scenario to the moment the participant hits “ARM”, or calls in his final site selection. (2) *Landing site score* (LSS) is the quality of the landing site based primarily on the roughness and slope. (3) *Participant Performance Score (PPS)* is defined as the participant’s ability to perform the LPD task and accounts for the LSS and the time to complete. These three dependent variables were measured after every run. (4) *Automation Trust* is based on a 5-point Likert scale regarding the participant’s overall trust of the automation, which consists of the suggested site rankings. A low rating (1) specified full trust in the automation system, never referring to the participant’s own judgment but relying solely on the automation ranking. Conversely, a high rating (5) denoted no trust in the automation system, never using the automation site ranking. (5) *Lighting Impact Opinion* is also based on a 5-point Likert scale and represents the participant’s opinion on the impact of lighting on his decision-making. A low rating (1) meant no impact, whereas a high rating (5) signified full impact, to the point where the participant changed his strategy based on the lighting quality. Automation trust and lighting impact opinion were measured at the end of each participant’s experiment. Lastly, the sixth variable was not directly measured but asked during the participant’s debriefing and involves the (6) participant’s strategy or decision-making procedure. The participants drew or enumerated a series of steps to complete the LPD task, including details such as method of site evaluation (direction, in order of suggested rank, etc) and interaction with the display.

3.2 Experimental Design

The major hypotheses of this experiment concern the impact of lunar lighting effects on pilot performance. Lunar lighting is treated as a within-subjects independent variable of two levels: poor and ideal, based on consultation with subject matter experts. The experiment also uses the nominal LPD initial trajectory conditions proposed by the ALHAT (vehicle is moving at 1km/s and is at 1km altitude). This trajectory provides an experiment baseline while representing a possible operation state for a moderate level of human control.

3.3 Participants

A minimum of 19 participants are needed for this study to achieve a power of 0.8 with an α value of 0.1. This number of participants is derived from the power calculations for a one-sample, one-tailed, paired t-test [1] and assumes standard deviation ($\sigma_{T_c}, \sigma_{m_f}, \sigma_{PPS}, \sigma_{LSS}$) values of 1 and absolute mean (μ) difference

[1] Using Russ Lenth’s java applet tool [20].

scores of 0.5. An α value of 0.1 was selected to reduce the possibility of a Type II error. For this experiment, the null hypothesis (H_0) states that the current system is sufficient to support the pilot during both lighting conditions. The alternative hypothesis (H_A) states that a new system must be adopted that accounts for two different lighting conditions. A Type II error is committed if the experiment results reject H_A, therefore keeping the old system, when lighting does, in fact, pose a significant impact on the dependent measures. Conversely, a Type I error is committed if the experiment results reject H_0 and adopt a new system when lighting did not have a significant impact on the dependent measures. Committing a Type I error would result in additional costs to the designer but would provide the pilot with all the tools he needs to complete LPD. Simplistically, the safety risk is low, since the tool exists on-board with the operator. However, a Type II error poses a greater safety risk to the pilot, as the necessary tools may not be present with the old system.

Six participants were tested during the pilot study of this experiment. Each participant was an engineering graduate student between 23-27 years of age. No participants had any flight experience, but all participants were familiar with vehicle dynamics and manned spaceflight. Each testing session was one hour long and consisted of initial briefing, training, and debriefing sections. The initial briefing introduced the LPD task, the simulator, and performance considerations that were explicitly included in the PPS (fuel and safety). To avoid bias, the participants were not told of the weighting on these parameters. The participants trained on ten scenarios similar to the actual runs. In each scenario, the participant was told he had 60s to complete the LPD task, otherwise the mission would abort. Every run began with an initial map of the landing area, with a resolution slightly less than the actual map. This initial map did not contain any highlighted hazards but did include lighting effects as seen from a top-down view. Participants were told they could study this initial map as much as they wanted and to start the scenario when they felt ready. The practice runs differed from the actual runs by including the sites' fuel and safety scores. Additionally, each site was magnified to illustrate the "actual" hazards within each landing footprint. The participants were told they could trust the hazard information, but similar to the actual performance of a sensor, there were occasional inaccuracies. All of the training materials are listed in Appendix.

After ten practice scenarios (total time approximately 15-30min depending on time spent on practice feedback), the participants were asked to complete two additional, actual runs. They were not given feedback on their performance. The experiment concluded with a debriefing, asking the participants to walk through the decision-making process of their two actual runs, to evaluate their Automation Trust and Opinion on the Lighting Impact, and to enumerate or sketch their procedure to complete LPD. The Appendix includes a full list of the debriefing questions.

3.4 Apparatus

3.4.1 Hardware

Since this pilot study was intended to collect initial information on the LPD task, a simple simulation was set up, with more focus on the LPD display. The LPD display is equipped with one main, top-down view of the expected landing terrain (including the lighting effects). This main window, as seen in the Appendix, contains highlighted hazard information overlayed on the terrain map. A set of yellow reference bars are included to denote the pixel length of 50m. Three sites are provided to the operator, listed in rank of automation preference. Site 1 is outlined in blue and corresponds to the blue button to the right of the main window; Site 2 and Site 3 are yellow and purple, respectively.

To make a site selection, the participant was asked to click on the corresponding site button with the mouse (wherein the gray circle turns yellow) and immediately click on the ARM button (the now yellow circle turns green). The simulation ends as soon as the participant clicks on the ARM button. In the case of an emergency, the participant could click on the red STOP button. No runs were aborted or canceled by the user.

This display was designed in Microsoft Powerpoint, with the button presses hyperlinked to other slides. These slides showed the pressed button "depressed" as to give visual indicator of what button was pressed. Additionally, audio was provided to the participant (*e.g.*, "Site 1 selected", "Initiating landing sequence"). XNote Stopwatch [12] was used to provide the 60s countdown clock and the timer. This external program was initiated by using the hotkeys G and O, respectively, which the participants were asked to type in to start the scenario. The timer was assigned to record the time whenever the right mouse button was clicked. This setup was placed on a standard PC using Windows 7 OS, but could have easily been placed on earlier operating systems, such as Windows XP.

3.4.2 Software

There were two major pieces of software used for this experiment: a map generation suite (Digital Elevation Map Maker (DEMMaker) and The Applied Physics Laboratory Navigator (APLNav) developed by Johns Hopkins University APL and a faux automated landing system. DEMMaker was used to create the lunar maps and the associated shading and the faux automation was used to find three landing sites and score the sites.

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 12 landing scenarios: an ideal and poor lighting of six landing maps around the South Pole. Each map was 180 $\times$ 180 m. Areas containing about 25$\%$ of hazardous slopes were used. Maps with approximately 25$\%$ completely shaded (*i.e.*, completely black) were used.

The lighting conditions presented in this experiment do not represent the true ideal and poor conditions at the South Pole. The ideal lighting, with maximum contrast, was created by feeding the APLNav software the coordinates of the Apollo 15 landing site, a location near the equator. This lighting condition does not occur naturally on the Moon. Additionally, poor lighting conditions on the Moon are actually under complete darkness. The geometry of the South Pole does not allow considerable light reflection, whereas the equatorial region permits so much that washout (where the terrain has low contrast due to too much light) is a real lighting instantiation. The realistic "poor South Pole" condition is not an appropriate scenario for this experiment as the participant would have no other option but to rely strict on sensor data. The date for ideal lighting was the 284$^{\text{th}}$ day in 2011; poor lighting occurred on the 306$^{\text{th}}$. An example of this is seen in Figure 3.

Once the maps were created, shaded, and analyzed, they were fed into the faux automated landing system. This faux algorithm was designed similar to the algorithm described by Cohanim and Collins [10], with several major differences. First, the faux algorithm used in this experiment notably did not take into account any information outside of the landing footprint itself. The reference footprint used was a 10m diameter lunar lander. Therefore, only the slope within the footprint was used in consideration for landing site selection. Second, two selection rules were used to narrow down all of the sites to three. A cost map was generated that held fuel consumption (assumed to be a perfect circle where the center point on the map invoked no fuel penalty) and slope at equal weighting. On the first pass, sites needed to match two criteria: (1) cannot reside within a 10m border of the map edge; and (2) percentage of hazardous pixels within the landing footprint must be less than a user-determined percentage of the pixel square area. This

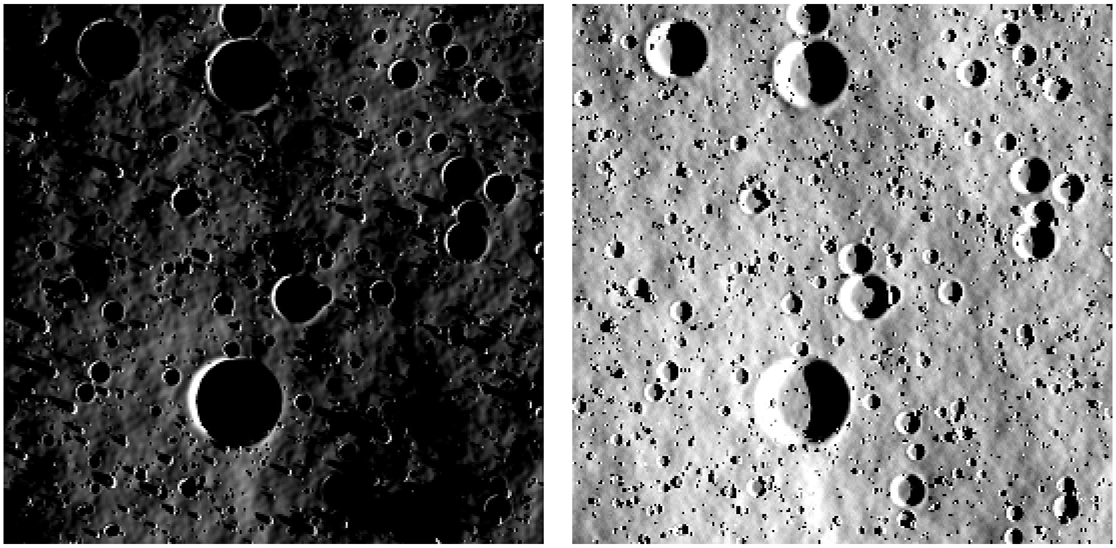


Figure 3: Difference between Poor and Ideal Lighting.

user-determined percentage was one of two fudge factors that could be adjusted, depending on whether three sites were found. On the second pass, the resulting sites from the first pass were examined based on rank (best to worst) and evaluated based on slope quality and overlap to other sites. The first three sites that contained less than a user-determined percentage of ‘poor slope’ and ’overlap’ became the three sites presented to the user. This second fudge factor could also be adjusted to account for more sites. This fudge factor was more sensitive than the first fudge factor. The two setting of the two factors became *de facto* indicators of scenario difficulty. Overall, an evaluation of twenty-plus maps resulted in six maps with the same fudge factors, roughly equal hazardous slope percentages and roughly equal dark shading. The landing locations are listed in Table 3

Table 3: Landing Locations.

Map #	Latitude	Longitude	Map #	Latitude	Longitude
1	-89.64	246.130	4	-89.64	251.303
2	-89.72	251.260	5	-89.72	251.270
3	-89.64	251.297	6	-89.72	296.276

An additional piece of code was used to score the sites. This scoring algorithm is similar to the faux automation but includes more information, similar to a theoretical, omniscient global performance evaluator. The fuel was no longer considered as a perfect circle radiating from the center map point, but instead, a series of elliptical contours (semi-major axis along the vertical) plotted at an eccentricity of 0.5. Sites reverse downrange (lower half of the landing map) received a penalty in addition to their distance score. The safety score now included roughness, calculated the number of pixels with a difference of 30cm from the average elevation with in the landing footprint. The safety (*i.e.*, LSS) and fuel scores ranged from 0 (perfect) to 1 (worst possible score). The PPS score was calculated using Eq. 1.

$$\frac{LSS + fuel_{score}(T_c/60)}{2} \tag{1}$$

This additional information did provide a different site ranking than that proposed by the faux algorithm. A list of the actual rankings are provided in Table 4.

Table 4: Comparison of Automated Rankings and Actual Rankings.

Map #	Auto-Site 1	Auto-Site 2	Auto-Site 3
1	Site 1	Site 2	Site 3
2	Site 1	Site 3	Site 2
3	Site 2	Site 3	Site 1
4	Site 3	Site 1	Site 2
5	Site 1	Site 3	Site 2
6	Site 2	Site 1	Site 3

4 Analysis of Results

This section discusses the results of the pilot study, including the overall characteristics of the participants' performance and the results of the statistical data analysis. All tests were run at a significance of $\alpha = 0.1$.

4.1 Overall Results

Six participants were tested in this experiment, with no participant needing to abort or cancel a run. Overall, the participants completed the LPD task in 13.656s ($\sigma_{T_c} = 6.545$), obtained an average LSS of 0.147 ($\sigma_{LSS} = 0.131$), and achieved an average PPS of 0.115 ($\sigma_{PPS} = 0.006$). The range of T_C is between 4.69 - 19.13s, with most participants completing the task in 9-15s (Figure 4). Boxplots of T_c, PPS, and LSS across the two lighting conditions are presented in Figure 5.

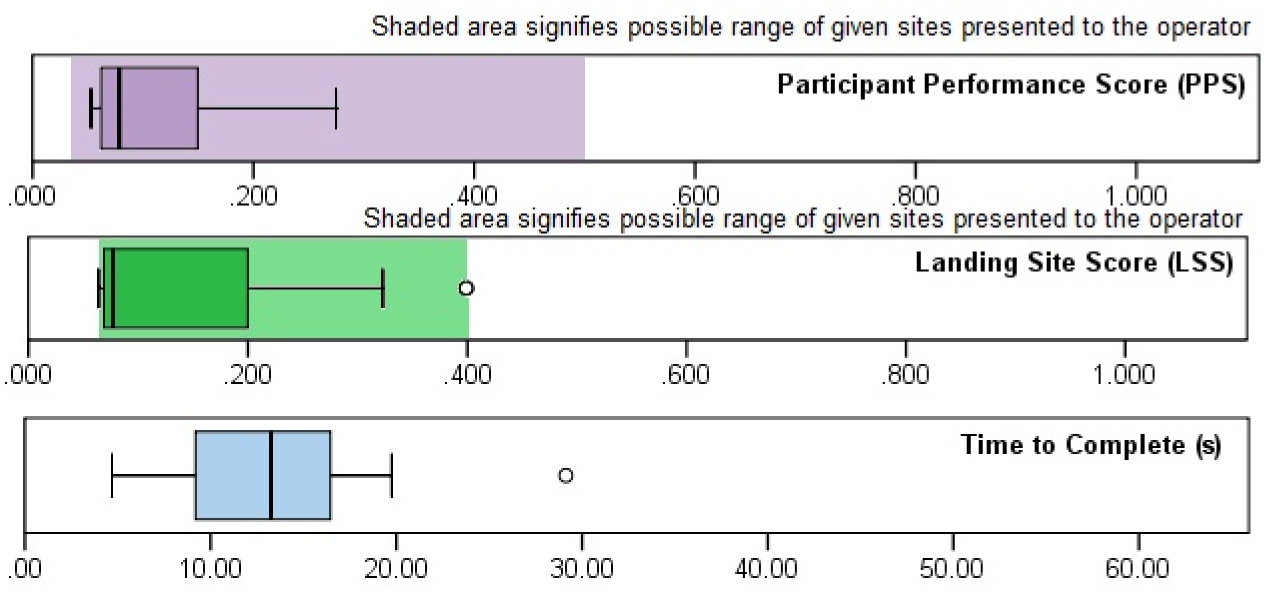


Figure 4: Total Range of Participant Performance Scores, Landing Site Scores, and Time to Complete.

The T_c and LSS were compared to determine if a significant correlation existed. No significant correlation exists, meaning that completing the task faster or slower did not help the participants pick better sites. Interestingly, T_c and PPS do not have a significant correlation. This result implies that T_c is not a dominating contributor to the PPS score, and the selection of a safe site is a larger driver.

The participants were also questioned regarding their opinion of the lighting impact and their trust in the automated landing system. Figure 6 illustrates a histogram of responses.

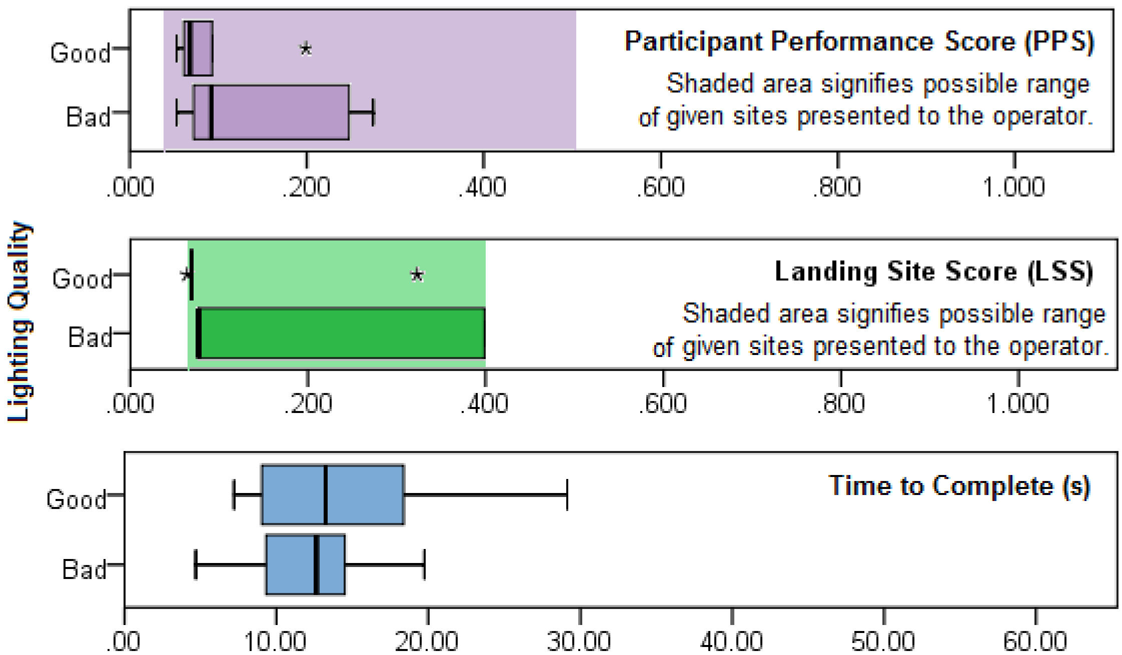


Figure 5: Range of Participant Performance Scores, Landing Site Scores, and Time to Complete across two lighting treatments.

As seen in Figure 7, participants generally agreed with the automation and selected Site 1. However, in the instance of Ideal lighting, this site was not the best, but the Site 3 presented to the participants was actually the best. Site 1 and Site 3 were very similar based on their respective LSS - the difference between the two sites was unlikely to have been detected by the participants. In the poor lighting scenario, a notable distinction between the presented Site 1 and Site 2 existed, with the presented Site 2 having about five times worst of an LSS compared to Site 1. Two participants did not recognize this difference.

4.2 Time to Complete

As discussed in Section 1, T_c is a critical measure due to its impact on fuel consumption. T_c was hypothesized to increase as the quality of the lighting decreased. The participants were assumed to require additional time to evaluate the areas of low visibility. The effects of lighting on T_c were tested using a Wilcoxon-Matched Pairs test. Lighting was not determined to have a significant effect on T_c.

4.3 Landing Site Selection

The LSS denotes the quality of the landing site and is ranked on an absolute scale. This absolute scale means that the quality of sites between landing areas can be compared on a one-to-one basis. The LSS was believed to improve with ideal lighting, as participants would be able to better evaluate the safety of the landing sites. The lighting effect was determined to be significant on the LSS using a Wilcoxon Matched-Pairs test with $Z = -1.370 (p = 0.0855)$. This result implies that the LSS in ideal lighting scenarios were better than the LSS in poor lighting scenarios. The boxplot of LSS with respect to lighting quality can be seen in Fig. 5. The means and standard deviations for the ideal and poor cases are $\mu_{LSS-ideal} = 0.184(\sigma_{LSS-ideal} = \pm 0.166)$ and $\mu_{LSS-poor} = 0.111(\sigma_{LSS-poor} = \pm 0.104)$, respectively. Lower values of LSS are better

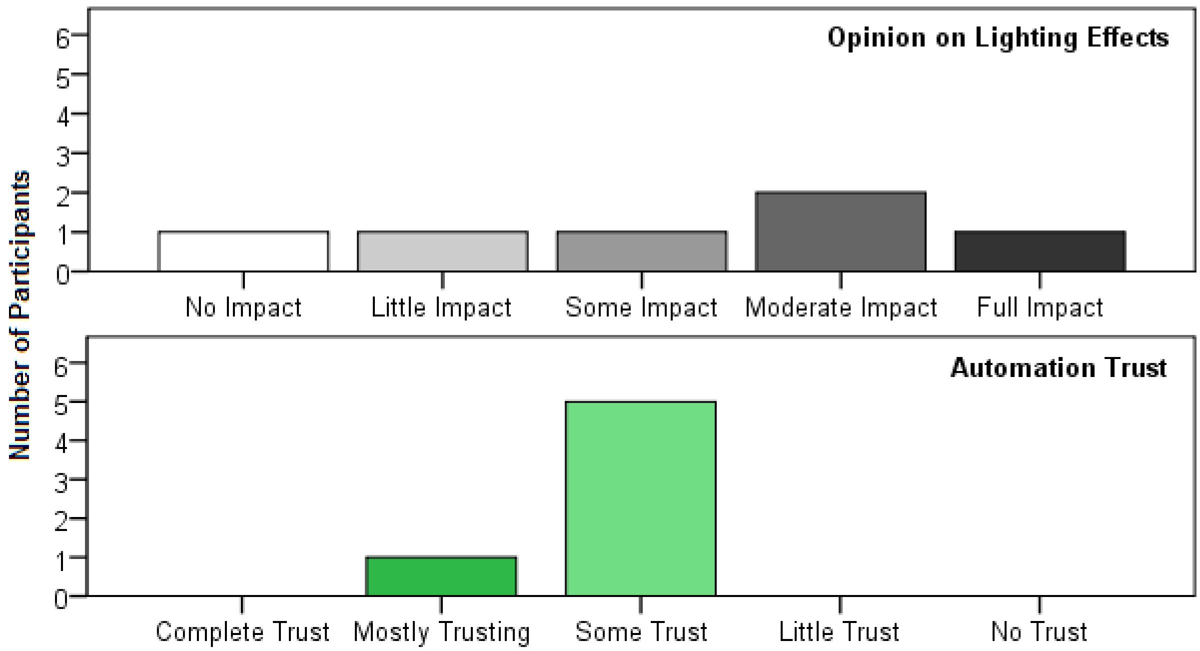


Figure 6: Histogram of Participants' Opinion on the Lighting Impact and Trust in the Automation.

than higher values, with 0 being a perfect score and 1 being the worst score possible.

4.4 Participant Performance Score

The PPS is a measure of the participant's performance. This dependent variable is calculated based on the LSS and T_c, on an absolute scale. Participant performances can be compared directly to each other without any transformation. The PPS was postulated to improve with ideal lighting, as the two factors contributing to PPS would similarly become worse with lighting. The lighting effect was determined to be significant on the PPS using a Wilcoxon Matched-Pairs test with $Z = -1.367(p = 0.086)$. The results of this test imply that the PPS in ideal lighting scenarios were better than the PPS in poor lighting scenarios. The boxplot of PPS with respect to lighting quality can be seen in Fig. 5. The means and standard deviations for the ideal and poor cases are $\mu_{PPS-ideal} = 0.090(\sigma_{PPS-ideal} = \pm 0.055$ and $\mu_{PPS-poor} = 0.139(\sigma_{PPS-poor} = \pm 0.097)$, respectively. Lower values of PPS are better than higher values, with 0 being a perfect score and 1 being the worst score possible.

4.5 Automation Trust

The automation trust metric was asked of the participants to determine whether the trust had a significant contribution to the other metrics (T_c, LSS, PPS). The participants were hypothesized to have worse scores in all metrics as the participant demonstrated less trust. A significant correlation does not exist between automation trust and T_c, LSS, or PPS.

4.6 Opinion on Lighting Impact

The participants were asked to rate their opinion of the lighting impact on their overall performance (in T_c, LSS, PPS). This measure was collected primarily to determine whether there was a notable difference

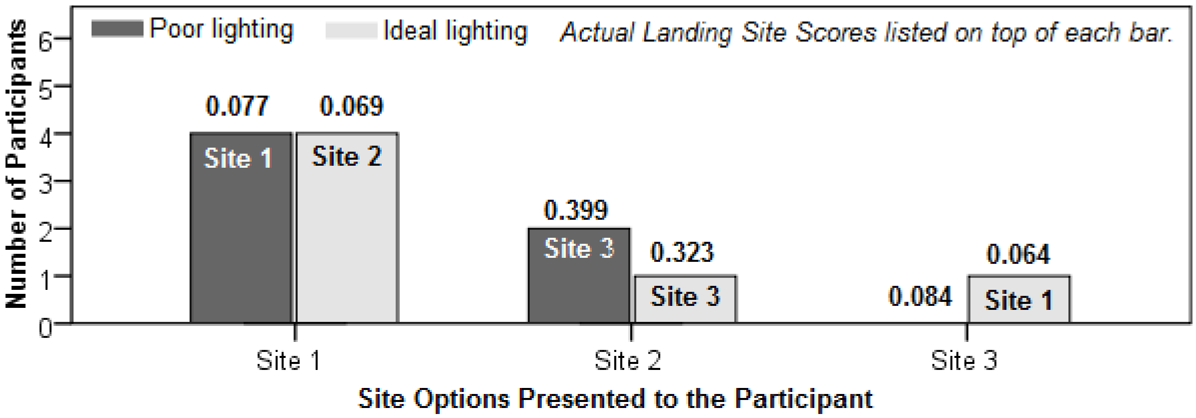


Figure 7: Histogram of Participants' Selection Compared to Actual Rank and Landing Site Score.

between perception and reality. The participants were expected to have worse scores as the participants noted less impact from lighting. Participants who neglected lighting were believed to not account for possible hidden craters or rocks within the landing area. The lighting impact opinion was deemed to have a significant effect on T_c, with a Spearman's rank-order correlation coefficient $\rho = 0.627, p = 0.008$. This correlation coefficient implies that as participants believe the lighting has a greater impact, the longer it takes them to complete the LPD task. As seen in Figure 5, the participants took more time to evaluate sites in the ideal lighting scenario, the inverse of the original hypothesis. There is no significant correlation between lighting opinion and LSS or PPS.

5 Discussion

Overall, the results of this experiment demonstrated that lunar lighting plays a significant effect on the LSS and the PPS, but not on the T_c. Participants do not perform the LPD task better, nor do they select better landing sites in more or less time. They are able to complete the task between 9-15s, which is consistent with previous literature. Participants generally chose better sites of the options presented to them. This site selection quality and brevity in selecting the sites meant the global PPS was within the better half of possible scores. The automation was an influence on participants' decision making, but was not the dominating influence. There is no significant correlation between the automation trust and T_c, LSS, and PPS. Participants had varying opinions on the effect of lighting, with one-third of the participants declaring the lighting to have a moderate impact. No significant correlation exists between lighting opinion and LSS or PPS, but a significant correlation exists between lighting opinion and T_c. These results have interesting implications. First, the participants did not believe that lighting had an effect on their performance, but lighting still had a significant impact on the quality of the sites chosen and the overall performance. Second, the participants who did take into account lighting took more time to analyze the maps, but this additional processing time did not actually have a significant impact on the time to complete. The implications of all of these results state that it may be of use to future designers to create two separate displays to account for the difference in lighting effects, as the current display may not be sufficient for both types of environments. However, further investigation is needed to specifically determine whether different displays are needed and where those changes should occur.

The participants were polled on their strategy during the debriefing. They were asked to recall their decision-making process during the two actual runs, to enumerate the qualities that factored into their decision making, and to present their task procedure to complete LPD, including their prescribed interactions with the display mechanism. Safety was a greater criterion than fuel for all six participants, with most participants defining perceived elevation difference (as noted by the shading) and size and density of highlighted hazards within the landing footprint. Fuel was used as a tiebreaker for sites with similar safety characteristics. This behavior and breakdown of considerations was expected and is consistent with previous studies. However, the proximity and location of major highlighted hazards was handled differently by the participants. Four participants generally chose sites that were far away from major highlighted hazards, citing concerns with landing short or long of the intended site. All four participants cited possible uncertainties with landing accuracy. However, two participants attempted to land near major highlighted hazards, opting for sites near clusterings of craters. Of the two, one participant actively sought to land near more populated groups of craters rather than large craters themselves. When pressed for justification for this behavior, the two participants independently stated the crater clusterings had more potential for scientific exploration and thus, these sites were preferable.

Many of the participants' strategies overlapped to form one global strategy. Figure 8 illustrates this global strategy and deviances from this central path. Half of the participants used the initial map to create mental models of the expected landing sites. These participants started from the center of the map and noted visible areas that did not have visible obstacles and not near craters. Some participants mentally ranked these areas and others actively set expectations for sites to show up in those areas. All participants, when starting the scenario, studied the highlighted hazardous areas. Those who used the initial map immediately eliminated sites not within their desired areas and set preference to those sites nearing or matching pre-scenario expectations. Further analysis showed that participants who studied the initial map on average performed the task in 13.363s ($\sigma = 2.983$s) compared to an average T_c of 13.948s (σ_8.333). However, a Wilcoxon-Matched Pairs test proves that this T_c difference is not significant. There is also no significant difference in LSS or PPS. All participants examined the landing sites for hazards within and outside of the circle, although the specific methodology ranged from looking center to outward; or in the order of the site rank; looking as far as 6 radius lengths beyond the site in a vertical than horizontal sweep, or looking simply for high densities of hazards. The fuel criteria was used as a deciding factor, with two participants further evaluating the science potential of the site. One participant recommended "double-checking" the selection after selecting site, and then hitting ARM; the other participants immediately hit ARM after selecting a site. No participants changed their minds after the initial site selection.

6 Conclusion

An experiment has been designed to determine the impact of lunar lighting on pilot performance during landing point designation, especially during moderate levels of human control. A pilot study consisting of six participants was conducted and the results analyzed to pose an initial approximation on actual study results. Overall, the participants completed the landing point designation task as fast as 4.69s and as long as 19.76s, with most participants completing the task between 9-15s. In the majority of all runs, the participants selected the top few sites presented to them. The lighting was varied at two levels: ideal and poor, and five main dependent measures were collected: time to complete, landing site score, participant performance score, opinion on lighting effect, and trust in automation. The participants selected higher-quality sites and performed the task better in the ideal lighting scenario than the poor lighting scenario. While participants did use more time during ideal lighting scenarios than poor lighting scenarios (contrary to the initial hypothesis), this change in task time is not significant. All of the participants followed one general strategy in completing the landing point designation task, with slight variations in search methodology and decision criteria. Half of the participants formed mental models to help them prepare for the actual hazard highlighting results, but this preparation was not significant in reducing the time to complete, landing site score, or participant

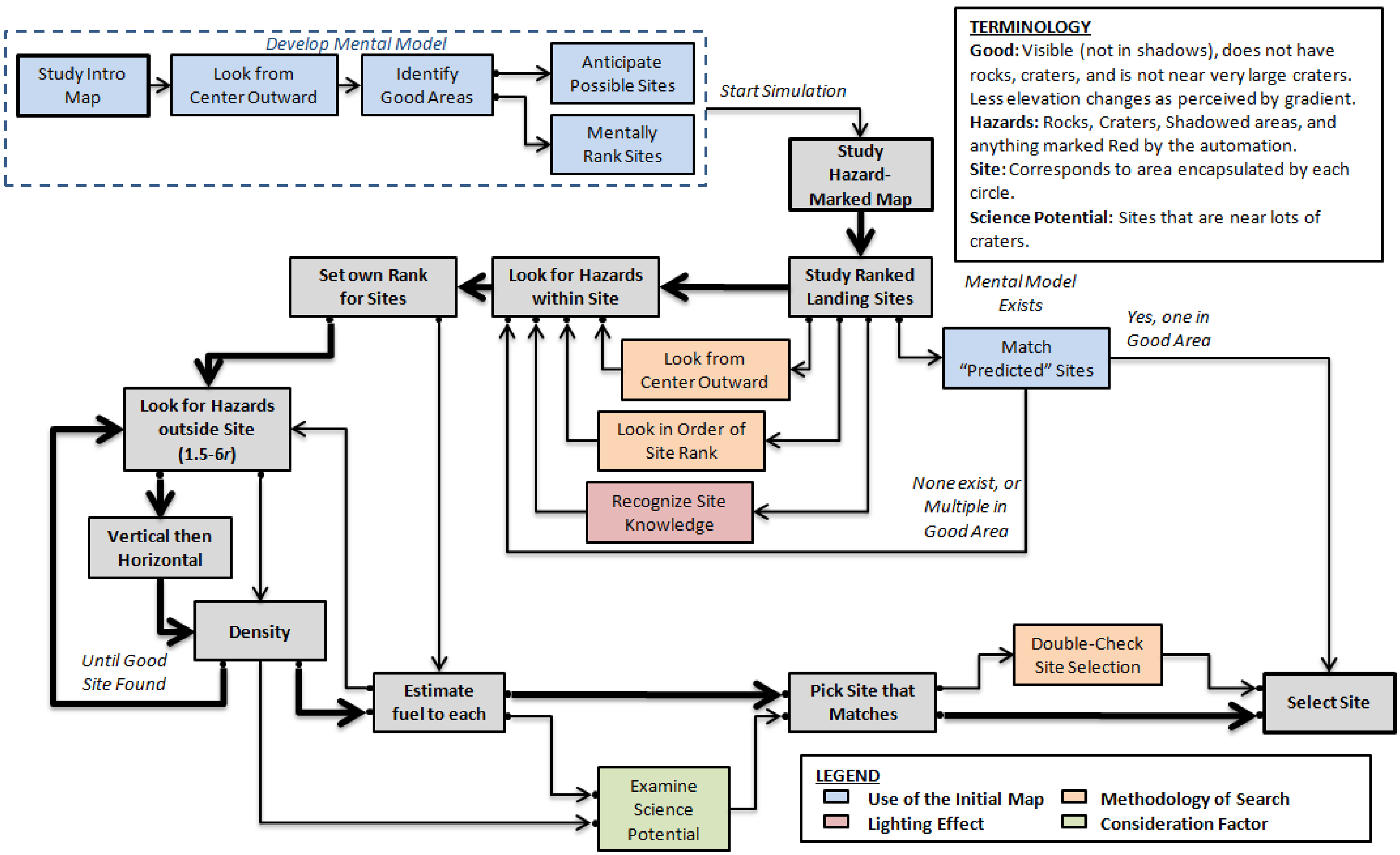


Figure 8: Flowchart of Participants' LPD Methodology.

performance score. Designing separate displays to account for the lighting differences may be of use to the participants, but additional research is necessary before a final conclusion can be drawn.

References

[1] National Aeronautics and Space Administration. Astronaut press conference. Press Release 60-276, National Aeronautics and Space Administration, Cape Canaveral, Florida, September 1960.

[2] Norman R. Augustine. Review of the u.s. human spaceflight plans committee. Technical report, U.S. Office of Science and Technology Policy, 2009.

[3] F. V. Bennett. Apollo lunar descent and ascent trajectories. Technical report, National Aeronautics and Space Administration, Houston, Texas, March 1970. Also presented at AIAA 8th Aerospace Sciences Meeting, January 1970.

[4] T. Brady, J. Schwartz, and T. Straube. Operational concept description for the autonomous landing and hazard avoidance technology (alhat). Technical Report ALHAT 2.0-002, National Aeronautics and Space Administration, 2006.

[5] Zarrin Chua, Todd Heino, and Robert Hirsh. Cockpit display documentation. Technical report, NASA Johnson Space Center, 2010.

[6] Zarrin K. Chua, Robert D. Braun, and Karen M. Feigh. Analysis of human-system interaction for landing point redesignation. Masters ae 8900 project, Georgia Institute of Technology, 2009.

[7] Zarrin K. Chua and Karen M. Feigh. Quantifying pilot performance during landing point redesignation for system design. AIAA Journal of Aerospace Computing, Information, and Communication. In Review., 2010.

[8] Zarrin K. Chua and Laura M. Major. Task modeling for lunar landing redesignation. In *AIAA InfoTech@Aerospace Conference*, 2009.

[9] Zarrin K. Chua, Laura M. Major, and Karen M. Feigh. Modeling cockpit interface usage during lunar landing redesignation. In *International Symposium of Aviation Psychology*, 2009.

[10] Babak Cohanim and Brian Collins. Landing point designation algorithm for lunar landing. *Journal of Spacecraft and Rockets*, 46(4):858–864, 2009.

[11] J.L. Davis, S.A. Striepe, R.W. Maddock, G.D. Hines, S. Paschall II, B.E. Cohanim, T. Fill, M.C. Johnson, R.H. Bishop, K.J. DeMars, R.R. Sostaric, and A.E. Johnson. Advances in post2 end-to-end descent and landing simulation for the alhat project. In *AIAA/AAS Astrodynamics Specialist Conference & Exhibit*, number AIAA 2008-6938, 2008.

[12] dnSoft Research Group. Free stopwatch and counter. Online, 2010.

[13] L.M. Forest, B.E. Cohanim, and T. Brady. Human interactive landing point redesignation for lunar landing. In *Proceedings of the IEEE Aerospace Conference*, Big Sky, MN, March 2008.

[14] L.M. Forest, R. Hirsh, and K. Duda. Results from the crew evaluation. Technical report, National Aeronautics and Space Administration, December 2008.

[15] L.M. Forest, L.J. Kessler, and M.L. Homer. Design of a human-interactive autonomous flight manager (afm) for crewed lunar landing. In *AIAA Infotech@Aerospace*, 2007.

[16] A.W. Holland and S.T. VanderArk. Task analysis of shuttle entry and landing activities. Technical report, NASA Johnson Space Center, March 1993. NASA-TM-104761.

[17] E. M. Jones and K. Glover. Apollo lunar surface journal. Online, July 2008. accessed 23 September 2008.

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

[19] David Knichel and James Albus. Levels of autonomy. Presentation, July 2003.

[20] Russ Lenth. Java applets for power and sample size. Online., August 2006.

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

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

[23] J.M. Needham. Human-automation interaction for lunar landing aimpoint redesignation. Master's thesis, Massachusetts Institute of Technology, 2008.

[24] Ron Noteborn. Apollo gnc software. Online., November 2001.

[25] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3):286–298, 2000.

[26] R.W. Proud, J.J. Hart, and R. B. Mrozinski. Methods for determining the level of autonomy to desgin into a human spaceflight vehicle: A function specific approach. In *Proceedings of the Performance Metrics for Intelligent Systems Workshop*, 2003.

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

[28] Philip J. Smith, C. Elaine McCoy, and Charles Layton. 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*, 27(3):360–371, May 1997.

[29] R. Sostaric and J. Rea. Powered descent trajectory guidance and some considerations for human lunar landing. In *AIAA Guidance, Navigation, and Control Conference and Exhibit*, number AIAA-2005-6287, 2005.

[30] Bradley A. Steinfeldt, Michael J. Grant, Daniel M. Matz, Robert D. Braun, and Gregg H. Barton. Guidance, navigation, and control technology system trades for mars pinpoint landing. In *AIAA Atmospheric Flight Mechanics Conference*, Honolulu, HI, August 2008. AIAA 2008-6216.

[31] Christopher D. Wickens. Commonsense statistics. *Ergonomics in Design*, pages 18–22, 1998.

[32] David Woods, Frank O'Brien, Tim Brandt, Lennie Waugh, Ken MacTaggart, Andrew Vignaux, Ian Roberts, Robin Wheeler, Rob McCray, and Mike Jetzer. Apollo flight journal. Online, August 2010.

Appendix: Debriefing Questions and Training Materials

1. Can you talk to me about your thought process for your first run?

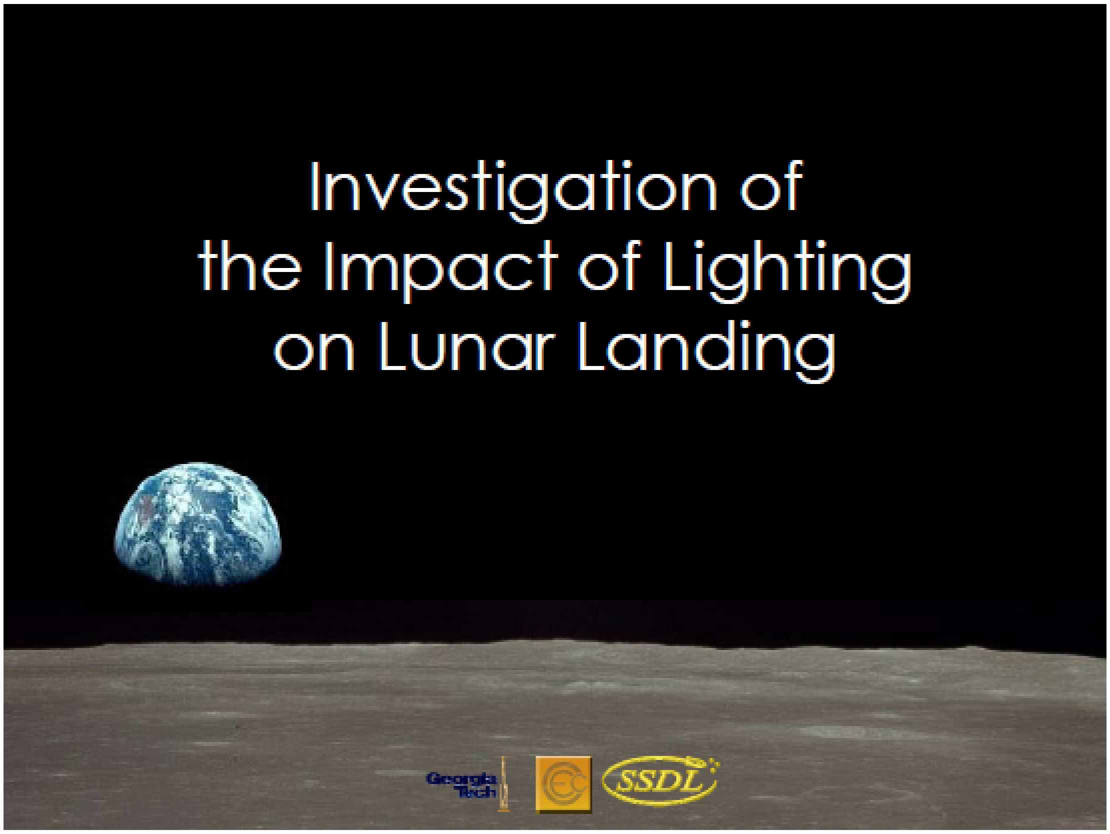
2. Can you talk to me about your thought process for your second run?

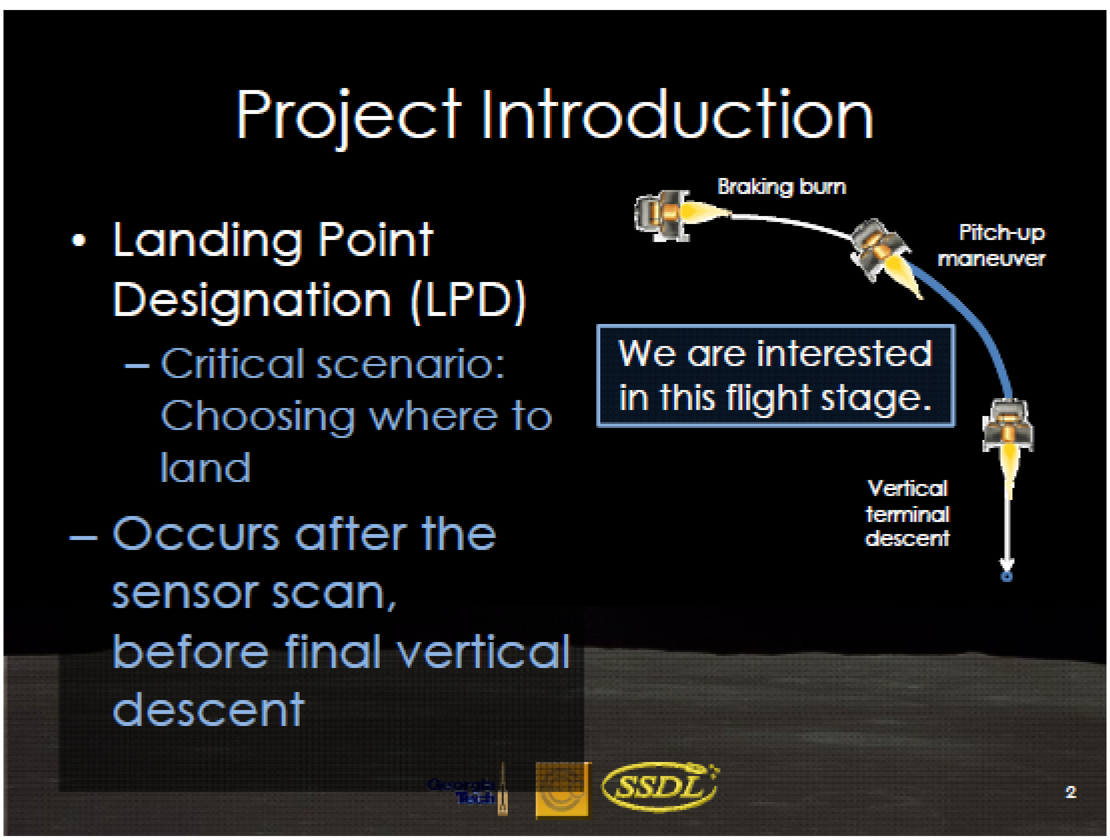
3. In summary, what decision attributes were you considering when selecting sites?

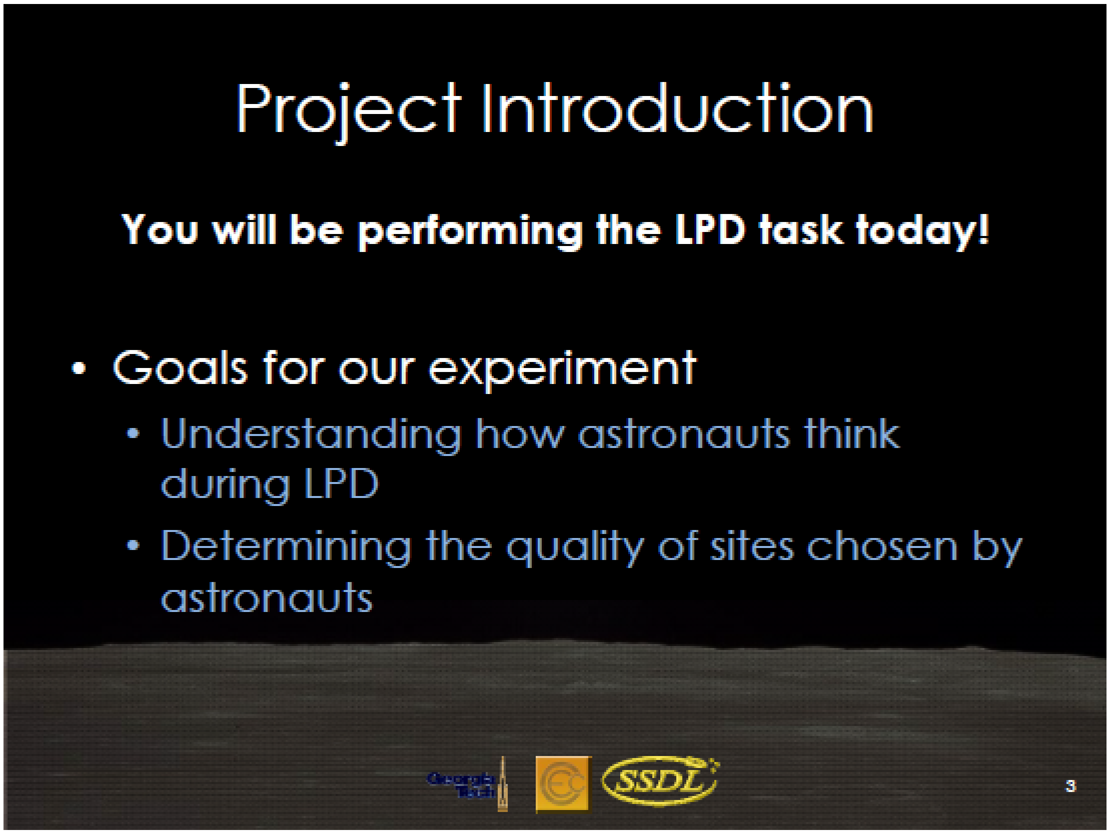
4. If you were training another person to complete this task, what universal strategy would you suggest to them? Can you draw me a flow chart or make me a bulleted list?

5. How must did you trust the automated landing system?
   1 - I trusted the landing system and always chose the suggested #1 site. I did not ever defer to my own judgment.
   2 -
   3 - I thought the landing system was accurate but occasionally questioned the rankings. I deferred to my judgment about half the time.
   4 -
   5 - I thought the landing system was not accurate at all and did not rely on the rankings. I relied solely on my judgment.

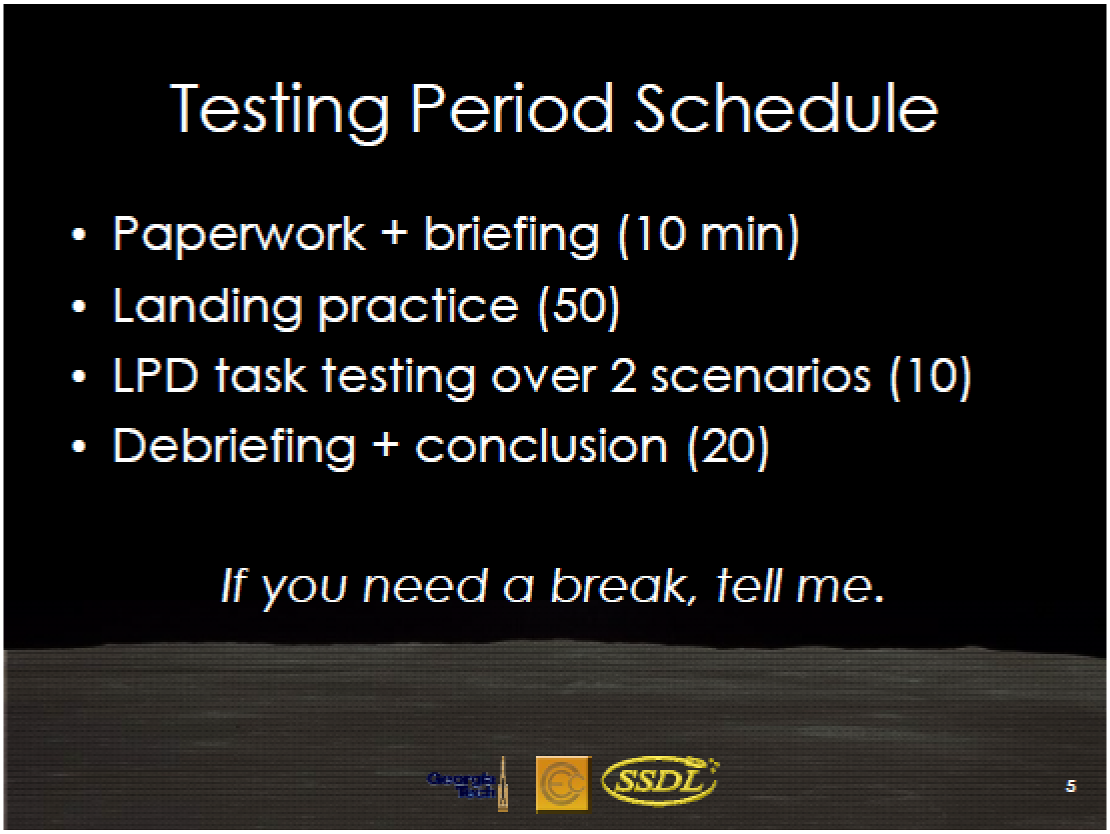
6. How much did lighting affect your performance?
   1 - More trust
   2 -
   3 -
   4 -
   5 - Less trust

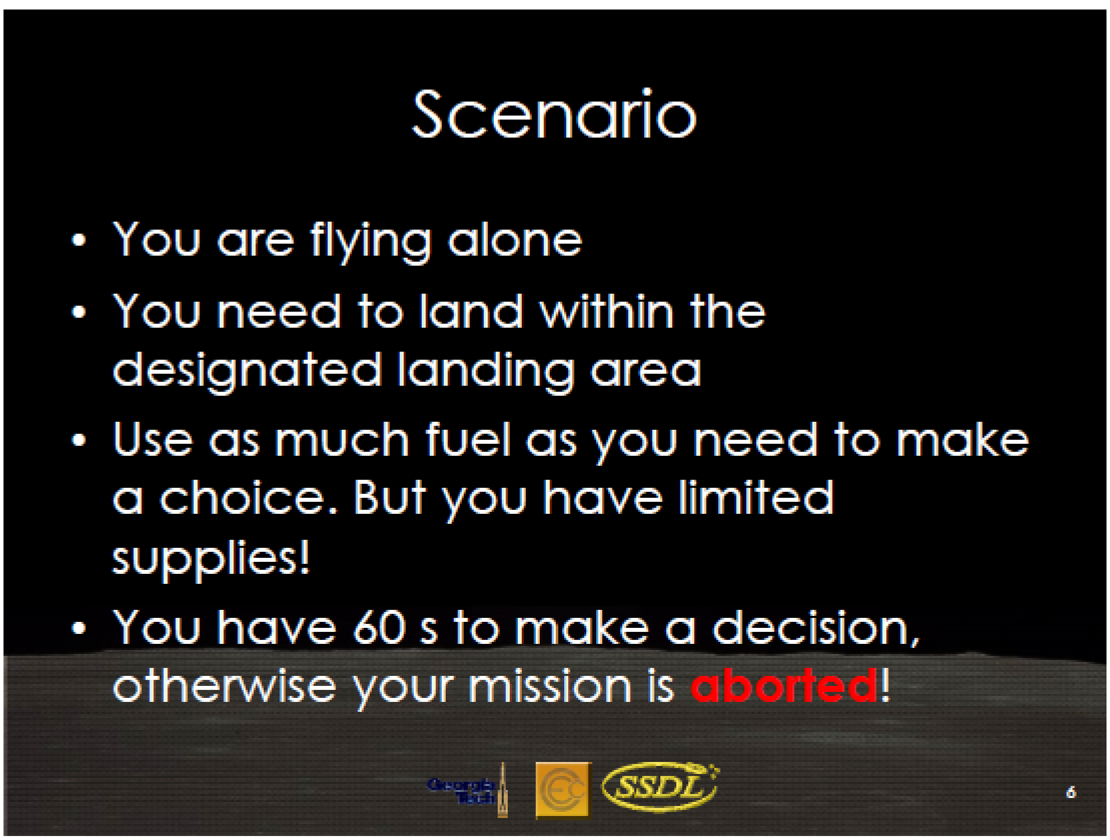

Investigation of
the Impact of Lighting
on Lunar Landing

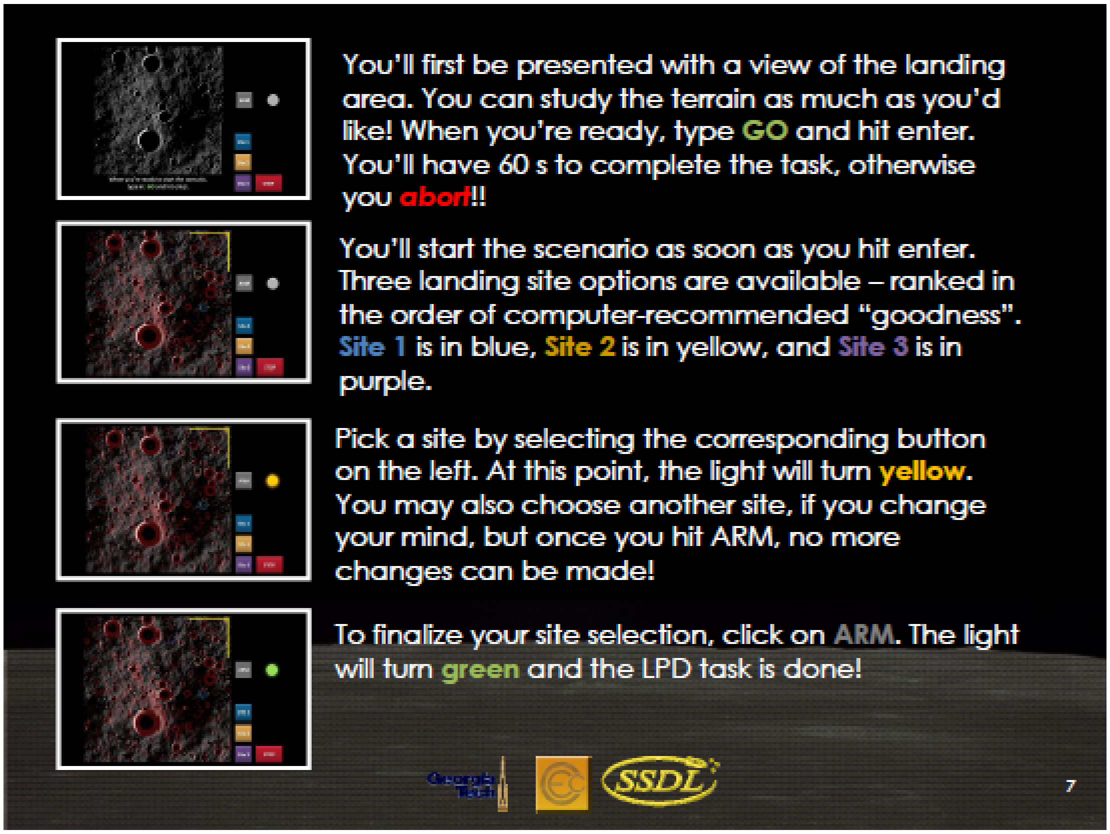

Project Introduction
• Landing Point Designation (LPD)
– Critical scenario: Choosing where to land
– Occurs after the sensor scan, before final vertical descent
Braking burn
Pitch-up maneuver
We are interested in this flight stage.
Vertical terminal descent
2

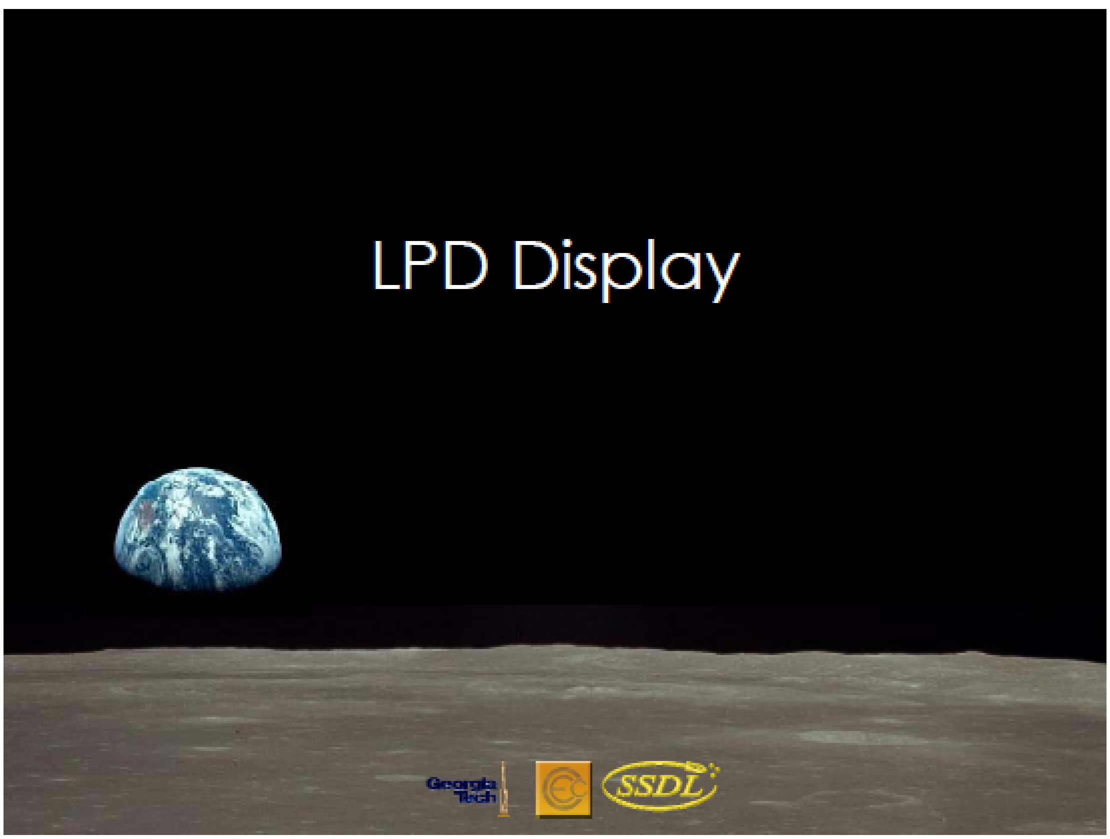

Project Introduction
You will be performing the LPD task today!
• Goals for our experiment
• Understanding how astronauts think during LPD
• Determining the quality of sites chosen by astronauts
SSDL
3

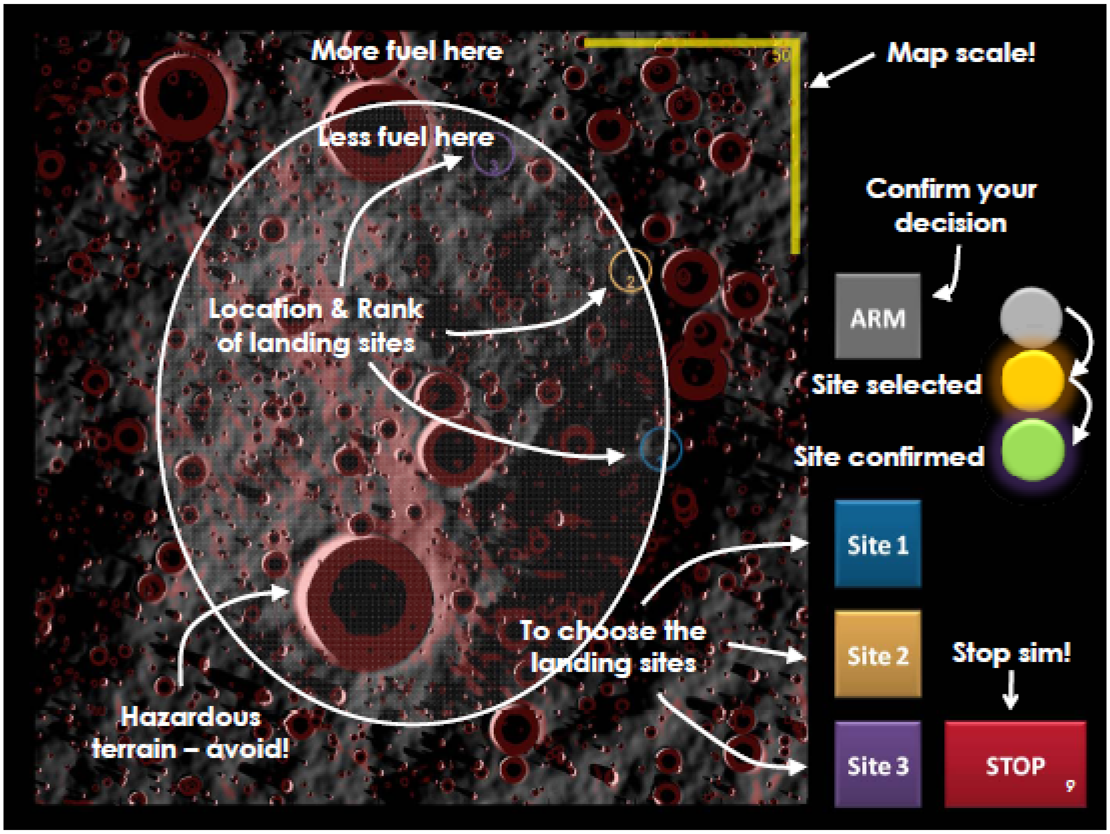

Participant Role
What you will be doing +
What you need to know
SSDL


Testing Period Schedule
• Paperwork + briefing (10 min)
• Landing practice (50)
• LPD task testing over 2 scenarios (10)
• Debriefing + conclusion (20)
If you need a break, tell me.
SSDL
5


Scenario
• You are flying alone
• You need to land within the designated landing area
• Use as much fuel as you need to make a choice. But you have limited supplies!
• You have 60 s to make a decision, otherwise your mission is aborted!
SSDL
6


You'll first be presented with a view of the landing area. You can study the terrain as much as you'd like! When you're ready, type GO and hit enter. You'll have 60 s to complete the task, otherwise you abort!!
You'll start the scenario as soon as you hit enter. Three landing site options are available – ranked in the order of computer-recommended "goodness". Site 1 is in blue, Site 2 is in yellow, and Site 3 is in purple.
Pick a site by selecting the corresponding button on the left. At this point, the light will turn yellow. You may also choose another site, if you change your mind, but once you hit ARM, no more changes can be made!
To finalize your site selection, click on ARM. The light will turn green and the LPD task is done!
SSDL
7


LPD Display
SSDL



Things to Consider

- Sites farther away require more fuel. You have limited supplies!
- Rocks can puncture, craters can leave you stuck, and slopes can cause your lander to tip over. Be careful where you land!
- The sensor data is accurate, but not all-knowing. Trust your judgment as well!

10