Examination of Human Performance during Lunar Landing

Zarrin K. Chua, Karen M. Feigh, Robert D. Braun
Georgia Institute of Technology
270 Ferst Drive
Atlanta, GA 30332
404-894-7783
{zarrin, karen.feigh}@gatech.edu, robert.braun@ae.gatech.edu

Abstract—Experimentally derived data was extrapolated to compare the lunar landing performance of human pilots to that of an automated landing system.[1 2] The results of this investigation are presented. Overall, the pilots performed equal to or better than the automated system in 18% of the relevant cases, but required more fuel. Pilot site selections were further investigated as a function of the time to complete. Each hypothetical case was compared to the automated system, across a range of performance criteria weighting distributions. This performance criteria is threefold – proximity to point of interest, safety of the site, and fuel consumed. In general, the pilots perform better than the automated system in terms of safety and proximity to points of interest criteria. However, as the priority of fuel conservation increases, the tradeoff between using an autonomous landing system versus a human-in-command system favors the automation, especially if the pilot is not able to make the proper decision within a performance criteria specific threshold.

TABLE OF CONTENTS

1. INTRODUCTION **1**
2. BACKGROUND **2**
3. EXPERIMENT DESIGN **2**
4. ANALYSIS OF EXPERIMENT RESULTS **4**
5. DISCUSSION **8**
6. FUTURE WORK **9**
7. CONCLUSION **9**
REFERENCES **9**
ACKNOWLEDGEMENTS **10**
BIOGRAPHY **10**

1. INTRODUCTION

The great achievements of human space exploration have generally been attained with some degree of automated assistance. This assistance, particularly in the areas of guidance, navigation and control has been manifested in advances in flight computers, control algorithms, and decision support systems. The distribution of work between these systems and the crew is focused on maximizing mission success and accounting for crew safety. The criteria for work allocation have been based on design heuristics, technical capability, crew preference and, in some instances, previous mission data. Extensive training prior to launch is used to further reinforce crew responsibilities and to test system capabilities. The human strength of adaptation and creativity provides the ability to compensate for technological deficiencies both known and unknown [1]. However, as crewed missions grow increasingly complex in both objectives and constraints, the guidelines historically used to determine the work allocation between crew and automation may not provide the most robust or optimal combinations. Instead a more sophisticated and quantifiable method is needed to determine the best allocation of work between human crews and automated systems. This method must account for and quantify both the implications for performance and mission robustness. Modeling and simulation of human-system interaction may be required to examine the overall design space and to make informed decisions.

The recent effort to land on the surface of the Moon is a prime example of the need for focused analysis of the tradeoffs between human-system interaction and the gains of leveraging the strengths of each member in this relationship. The return to the Moon will require the next generation lunar lander to touchdown in regions far more hazardous than that experienced during the Apollo missions. These regions are typically poorly lit with distinctive terrain features. The crew will likely require some level of automated assistance in order to achieve a safe and precise landing [2].

The landing task is comprised of multiple phases, such as a major vehicle braking burn, navigation sensor calibration, landing point redesignation (LPR), and a terminal vertical descent. During the LPR task, the crew selects a final touchdown location that meets fuel consumption, safety, and proximity to points of interest criteria. This task, and its inherent reliance on the crew to finalize the touchdown point, generally requires a non-fuel-optimal trajectory to orient the spacecraft appropriately to view the landing zone and to allow the crew time to make a decision. Prolonged decision-making is costly as there is a high correspondence between time and fuel consumption. Questions about the tradeoff between fully automating the LPR task versus allowing the crew to remain in-command are still unresolved.

This paper presents a theoretical investigation into the tradeoffs between human and automated performance

[1] 978-1-4244-3888-4/10/$25.00 ©2010 IEEE
[2] IEEEAC paper #1150, Version 7, Updated 12/29/09

during LPR across possible figures of merit for mission success. The human performance data used here is based on the results of an experimental study that was conducted to determine the speed and performance of human-pilots during the LPR task [3]. Special attention was given to the task completion time and the quality of the site selection. This paper will summarize the experimental setup necessary to examine the LPR task, the experimental findings, and the comparison of human performance to analogous automation.

2. BACKGROUND

With the increased expectations for next generation lunar landers, recent studies related to crewed lunar landing have focused on the development of support systems (i.e., displays and landing algorithms). One such effort is the development of an Autonomous Flight Manager (AFM) by NASA's Autonomous Landing and Hazard Avoidance Technology (ALHAT) team, led by NASA Johnson Space Center. The AFM is analogous to a flight management system – providing guidance, navigation, and control cues [4], monitoring system health [5], and interacting with the crew, including prompts for supervisory commands [6]. With respect to the LPR task, the AFM serves two purposes: 1) processing raw sensor data into a form comprehensible to the crew, and 2) from this sensor data, suggesting alternative landing sites to the *a priori* baseline site.

Until this point, the only source of terrain information available to the crew is an unaugmented window or camera view. After AFM processing, the crew has the window or camera view and the results of the LIDAR scan as additional sources of terrain information. The crew evaluates the alternative landing sites, finds a site that satisfies its specified criteria (e.g., safety, required fuel, or nearness to point of interest), and designates the final landing site, which concludes the LPR task. The LPR task must be completed quickly, as the trajectory required to enable LPR is typically not fuel-optimal. Currently, the LPR task is expected to occur during the Powered Descent Phase. Just prior to LPR, the vehicle performs a pitch-up maneuver, placing the vehicle in an orientation suitable for LIght Detection and Ranging (LIDAR) sensor operation [4]. This maneuver is expected to occur at approximately 1 km in altitude, at a velocity of 100 m/s (nominal trajectory) [7].

As with the Apollo missions, the landing trajectory is designed in a way to provide the crew a visual of the landing area [8]. However, conditions at desired landing areas, such as the far side of the Moon, may impede the crew's ability to acquire terrain information unaided. Preliminary analyses estimate 30s are needed for LPR task completion [9]. In this period of time, the astronauts must absorb information from the AFM and window or camera view; perform tradeoffs of safety, fuel consumption, and proximity to the Point Of Interest (POI), and select a final landing site. Likewise, the crew must adapt to any unanticipated terrain features.

3. EXPERIMENT DESIGN

Sixteen lunar landing scenarios were developed from combinations of three independent variables: *Points Of Interest (POI)* which are one or two landing sites that reflect the purpose of the mission; *terrain expectancy (ε)* denoting whether the pre-launch lunar terrain training matches or does not match the actual lunar terrain; and *identifiable terrain markers (ITM)* which are one, two, three or four clusters or formations of hazards. This experiment blocked the scenarios based on ITMs – participants saw a "high" and a "low" density of ITMs, creating two groups of ITM (1,3) and ITM (2,4). Thus, of the full sixteen scenarios from a 2 × 2 × 4 full factorial design of experiment, each individual participant experienced eight scenarios. The order of the runs within and between subjects was balanced to reduce any potential bias in run order. This experiment collected several dependent measures: time to complete, quality of landing site selection, task strategy, pilot workload, situation awareness, and display effectiveness.

Twenty pilots participated in this experiment, representing a wide variety of flight experience and pilot training. Participation in the experiment was limited to individuals holding a Private Pilot License (PPL) and at least 80 hours of flying experience. This stipulation ensured enough familiarity with standard aircraft landing procedures and the process of selecting a landing site without limiting the number of samples for statistical accuracy. Twelve pilots were Visual Flight Rules (VFR) rated, seven were also Instrument Flight Rules (IFR) rated, and one pilot had an additional Commercial Pilot License (CPL). The pilots have flown single- and multi-engine aircraft both for personal and commercial use. The mean was 277 h for flying (σ = 307 h). No military pilots participated and only one pilot had experience flying helicopters. The majority of the participants were less than thirty years old. The pilots, unknown to them, were randomly separated into two groups for ITM frequency blocking. There were eight participants (six VFR, two IFR) in the ITM (2,4) group and twelve (six VFR, five IFR, one CPL) participants in the ITM (1,3) group. Fig. 1 illustrates the distribution of the hours of flight experience. The flight experience mean was 181 h in the (2,4) group ($\sigma_{(2,4)}$ = 140.5 h) and 340 h in the (1,3) group ($\sigma_{(1,3)}$ = 373.1 h). The discrepancy in flight hours and inequality of VFR/IFR pilots was not determined until after the experiment and as such, flight hours and pilot certification are included as covariates in the data analysis.

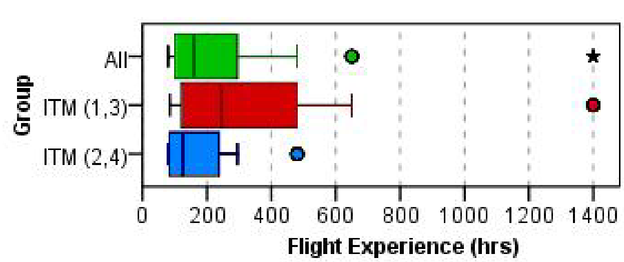


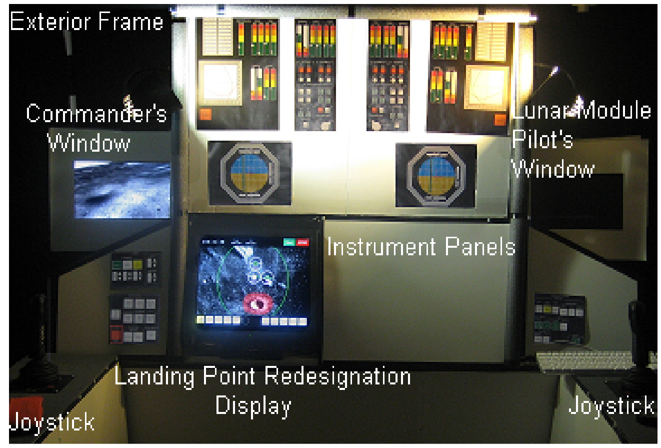
Figure 1 – Distribution of Flight Experience (h).

Likewise, the experiment setup was intended to have an equal distribution of pilots for both groups, but there were instances of inability to train to proficiency and a simulation failure, which led to exclusion of that data in the analysis. Each successful testing session lasted two hours. The initial briefing introduced the LPR task and the simulator. The pilots practiced the LPR task for 45 min in the simulator, where they received feedback on their performance. This exercise allowed the pilots to become comfortable with the simulator and to formulate strategies. The testing session was comprised of eight runs of pre-selected specified landing scenarios.

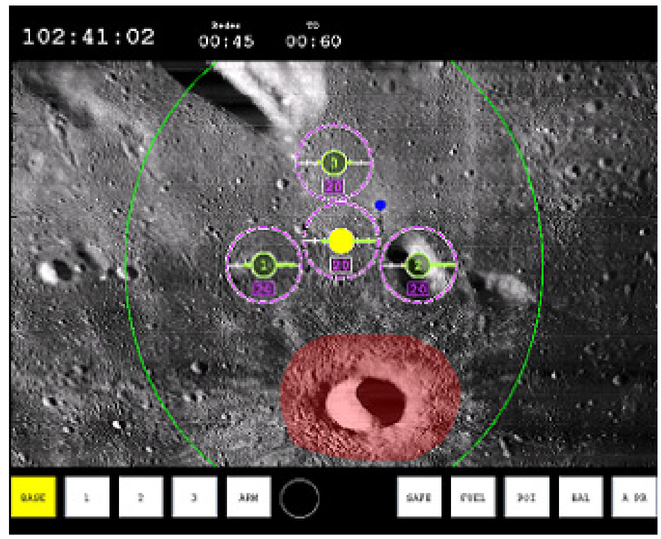
The experiment utilized a simulated lunar lander module, which included an LPR reference display incorporating an AFM guidance algorithm and an out the window view. A full description of the simulation facility, illustrated in Fig. 2, can be found in [9, 10], and illustrated in Fig. 2. This display design was chosen based because it provided a publicly available symbology template for use in this investigation. The software used in this study consists of two major components: dynamic elements used to enhance the lunar lander mockup hardware and a Pseudo-AFM LPR algorithM (PALM) that emulates the AFM algorithm of alternative landing site selection [11, 12]. An out-the-window perspective was provided by using the EagleLander3D [13] software. This perspective showed terrain as seen during vehicle pitch-up, approach, and terminal descent. The pilots remarked the out-the-window display added to the realism of the simulation.

As previously discussed, the lander is expected to be equipped with an AFM that offers alternative landing sites based on an objective function that is set by the crew. The PALM developed for this investigation takes an input package (lunar satellite photography map, hazard identification, POI location), scans the map for alternative landing sites, and outputs sites based on: safety, fuel efficiency, and proximity to POI. Five objective functions are calculated – one for each of the three individual metrics, and two others in which an equally balanced or *a priori* weighting of these three metrics is computed. This input package is read, and the landing area is converted to a matrix, with each cell containing a value from 0 to 255 (grayscale). The PALM treats this matrix as a LIDAR sensor scan, with each cell location corresponding to a geographical position and cell magnitude relating to an altitude. LIDAR measurement error is not modeled. The PALM measures a Euclidean vector difference between each non-hazardous cell, and the POI and examines the area within the landing footprint for terrain characteristics and fuel consumption requirements. Information on the calculation of slope, roughness, and fuel consumption can be found in [3].

Once these terrain characteristics are computed, the PALM sorts the sites based on each of the five objective functions. Logic is included in the algorithm to ensure that unique sites are recommended - no landing site overlaps another within



(a) Full Mock Lunar Lander View.



(b) Landing Point Redesignation Display. *The objective function buttons are located in the lower right corner.*

Figure 2 – Lunar Lander Simulation Environment.

the same objective function. The output map image for each objective function contains hazardous area highlights, the point(s) of interest, the baseline point and three alternative sites, and symbols for the relative goodness of slope and roughness of the expected landing area. These maps are generated prior to the experiment, and the map display corresponds to the actions of the user simulating a real-time calculation of alternative landing sites without the computational cost or increased risk of simulation failure.

The PALM also computes the pilot performance score (PPS) and rank (PPR). The PPS is calculated using Eq. 1 and the PPR is based on the rank of the selected landing site relative to the sites available. Eq. 1 is based on Voltaire's concept of "perfect is the enemy of good". Although the mission objective is to place the lander in an area free of major hazards and with preferable terrain characteristics (flat and level), there exists a region within each metric that constitutes sufficient performance. This region must be factored into the performance formula to account for more realistic figures of merit.

$$P_{score} = w_f f_w F_{LA} + w_p D_{POI} + w_s (S_{LA} + R_{LA} + D_h)/3$$

$$f_w = 1 \text{ - time to decision/total time for LPR} \quad (1)$$

where D_{POI} and D_h are the distances from the POI and hazards, F_{LA}, S_{LA}, R_{LA}, are raw scores for fuel consumption, slope, and roughness of the landing site, and w_f, w_p, and w_s are the weighting distributions for fuel consumption, proximity to POI, and safety. The sum of these weights must be equal to 1. The element of time is also introduced in this performance formula as a contributor to fuel consumed. Eq. 1 was also modified to evaluate the static properties (safety, proximity to POI) of the landing site itself, Eq. 2, by eliminating fuel as a performance measure.

$$P_{site} = (S_{LA} + R_{LA} + D_h)/3 + D_{POI} \quad (2)$$

4. Analysis of Experiment Results

This section summarizes the results of the experiment, including global participant performance and the effects of the independent variables. Unless otherwise stated, the statistical analysis was performed using both parametric (repeated measures ANalysis Of VAriance) and non-parametric (Spearman's correlation (ρ), Kendall Tau correlation (τ), or Friedman's test) tests as appropriate. Significance for all tests was set at $\alpha = 0.05$. For the full set of results and analysis readers are referred to Chua and Feigh [3].

Overall Results

Twenty pilots participated in the evaluation. Nineteen completed all eight runs, and the twentieth pilot completed five of the eight runs (due to a simulation failure), for a total of 157 cases across the full design of experiments. All pilots completed the task within the 45s allotted; no pilot aborted a run. On average, the LPR task was completed in 20.39s (σ = 9.05s) [4.08, 41.53s]. In 54% of the cases, the pilot chose one of the top site selection rankings, whereas the pilot made a poor selection in 7% of the cases. While these poorer site selections resulted in feasible landing locations, better sites were available at the selection time. Fig. 3 illustrates the distribution of top pilot certification (TPC) by final pilot performance ranking (PPR).

In general, the pilots preferred sites which were affiliated with an objective function over those sites associated with the *a priori* set, as seen in Fig. 4. The pre-launch designated baseline site, which could be selected at any time during the LPR task, was included under the *a priori* objective function, as the criterion dictated the location of the baseline site. The pilots were told that the vehicle would default to landing at the baseline site. The pilots were instructed to select an alternate or confirm the default. Should a landing

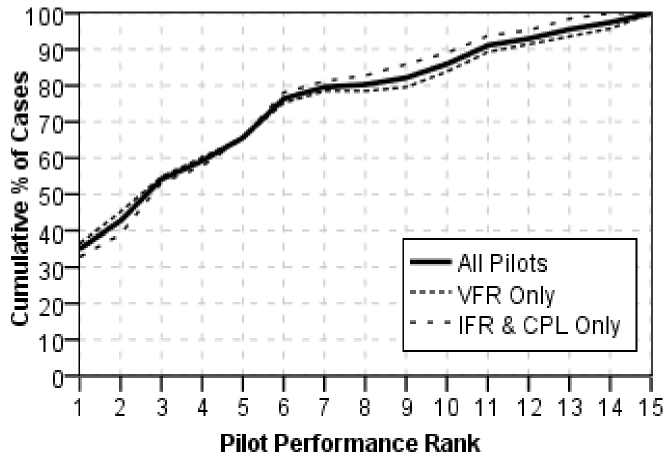


Figure 3 – Distribution of Final Pilot Performance Ranking.

site not be selected within the LPR task time, the lunar mission would abort.

Time to Complete

The LPR task completion time, T_C, is a critical value, impacting elements of lunar landing such as the fuel consumption and the design of the descent trajectory. This modeling hypothesized that an increase in the number of factors would result in additional pilot information processing time. In particular, the terrain expectancy factor, (ε), was expected to cause expert pilots to re-orient to the scenario, as described by Klein [14]. A repeated-measures ANOVA was utilized to test these hypotheses. POI had no significant effect on T_C for either the ITM (1,3) or the ITM (2,4) groups. Similarly, ε did not significantly affect T_C for either ITM (1,3) or ITM (2,4). The number of ITMs does not significantly affect T_C for ITM (1,3) or ITM (2,4). The interaction between ε and ITMs had significant effect on T_C for the ITM (2,4) group only, $F(1; 3) = 10.349$, $p = 0.049$. The pilots performed the LPR task faster as the number of ITMs increased in cases of unexpected terrain.

Pilot Performance

Pilot performance was calculated as both a continuous variable (pilot performance score, PPS) and as an ordinal variable (pilot performance ranking, PPR). The pilot performance was calculated from a weighted sum (equivalent to PPS) as described in Section 3, Eq. 1. This analysis uses both forms of the variable. Analyses regarding PPS were performed using repeated-measures ANOVA, while the Kendall Tau correlation was implemented to examine the relationship between PPR and other factors.

First, the PPR was hypothesized to become worse with the number of POIs and ITMs, and especially in instances where the terrain was unexpected. The effect of POI was not significant on PPR for either ITM group. Similarly, the effect of terrain expectancy was not significant for either

ITM group. The effect of the number of ITMs on POI is marginally significant for the ITM (1,3) group, $F(1; 6) = 5.926$, $p = 0.051$ but is not significant for the ITM (2,4) group. Thus, none of the independent variables included in this investigation were found to have significant effects on PPR.

Extrapolation of Results

While this experiment was not designed to investigate or compare the effectiveness of human control versus automated control, given the framework of the simulated task, the experimental data can be extrapolated to provide initial insight on human capabilities. Specifically, we have used the data collected in this experiment to speculate about the likely capabilities of an astronaut crew and their impact on system performance during LPR compared to an automatic landing system. However, several assumptions and definitions must be introduced to perform this analysis.

First, we assume that there exists some automated landing system capable of autonomously guiding the lander to any point on the lunar surface. Next, we assume for comparison that a vehicle equipped with such automation flies the same trajectory as that used in this experiment, which is notionally based on the ALHAT trajectory [7], thereby simplifying the comparison of fuel consumption to a dependence on task completion time. We acknowledge that this assumption is not likely to hold, as a fully automated vehicle would likely fly a more fuel optimal trajectory. However, lunar fuel optimal trajectories do not generally permit crew viewing of the landing site. Using this reference trajectory is useful for this level of comparison as the actual trajectory planned is still under development [16]. There may also be situations where the crew is incapacitated and a fully automated landing sequence is required, or adaptable automation is employed on the same flight, alternating or limiting crew or system command. We further assume that the choice of landing site is static for the automation and corresponds to the *a priori* site used in the experiment. Automated control therefore implies selection of the baseline site, completed in zero seconds.

The pilots' site selections and times to complete are assumed to be a conservative representation of astronaut behavior, or human control, during lunar landing. This definition is best analogized to a Monte Carlo analysis, where the pilots' behaviors in the experiment are assumed to be the behavior of one astronaut; the differences in landing scenarios and pilot experience is similar to the uncertainties associated with the inputs to a system; and the landing site selection and completion time are the result of performing a lunar landing under the prescribed inputs. As such, the actual site selection data is a sampling of the design space between the experiment independent variables and covariates (POI, ITM, ε, TPC, flight experience) and dependents (site selection, completion time). For purposes of this analysis, these site selections are further assumed to be pilots' best selections within the context of the landing scenario and the time to complete. This assumption suggests a layer of time homogeneity – should the pilots repeat the experiment, then the site selection is constant and would not improve or worsen. However, the site selection itself means little without assigning some definition of quality.

As defined for this analysis, the human controlled site selection will require more time and thus, more fuel consumption. Preliminary modeling shows that astronauts require 12-28 seconds to absorb the information and make the LPR decisions. However, an astronaut can potentially make better decisions than the automated system as to where to land. This ability is because of the potential for the astronauts to take advantage of more accurate information or information not available to the automation (i.e., algorithm failure). To ensure a fair comparison, only data from scenarios where the *a priori* site was mid-ranked (5-10^{th} out of a possible 15) were used. Fifty-six runs fall into this category. Additionally, a definition of performance quality is used to compare human control and automated control. This analysis uses Eq. 1, with arbitrarily set weights for the three metrics of interest, fuel consumption, proximity to POI, and safety, summing to 1. A 10% uncertainty is applied to the comparison of PPS and Automated Performance Score (APS) – if the PPS is within 10% of the APS score, then the PPS score is considered equivalent. Under this weight distribution and the three prescribed definitions of the automated system, the reference human behavior, and performance quality, the experiment results show that the pilots performed better than or equal to the automated system in 18% and worse than a basic automatic landing system in 82% of the selected cases. This percentage is heavily influenced by decision time because of the correlation with fuel burn. However, faster decisions do not necessarily imply better site selection.

To further explore this tradeoff, the distribution of safety and proximity to POI scores (using Eq. 2) with respect to fuel consumption (equivalent to task completion time) was plotted. As seen in Fig. 5, pilots' time to complete varies over the 45s of allotted LPR task time. Additionally, the pilots' ability to select quality sites is varied across time to complete. The Spearman correlation is non-significant at $\rho = -0.112$, indicating that as the pilots' task completion time increases; the ability to select a safe site near the POI is diminished. However, when the penalty for additional decision time is removed, the pilots consistently outperform the automated system with regard to selecting quality sites. In 63% of the cases, the pilots recognized and selected sites that were superior to the baseline site. This trend implies that humans are capable of selecting landing sites that minimize landing risk while increasing mission efficiency.

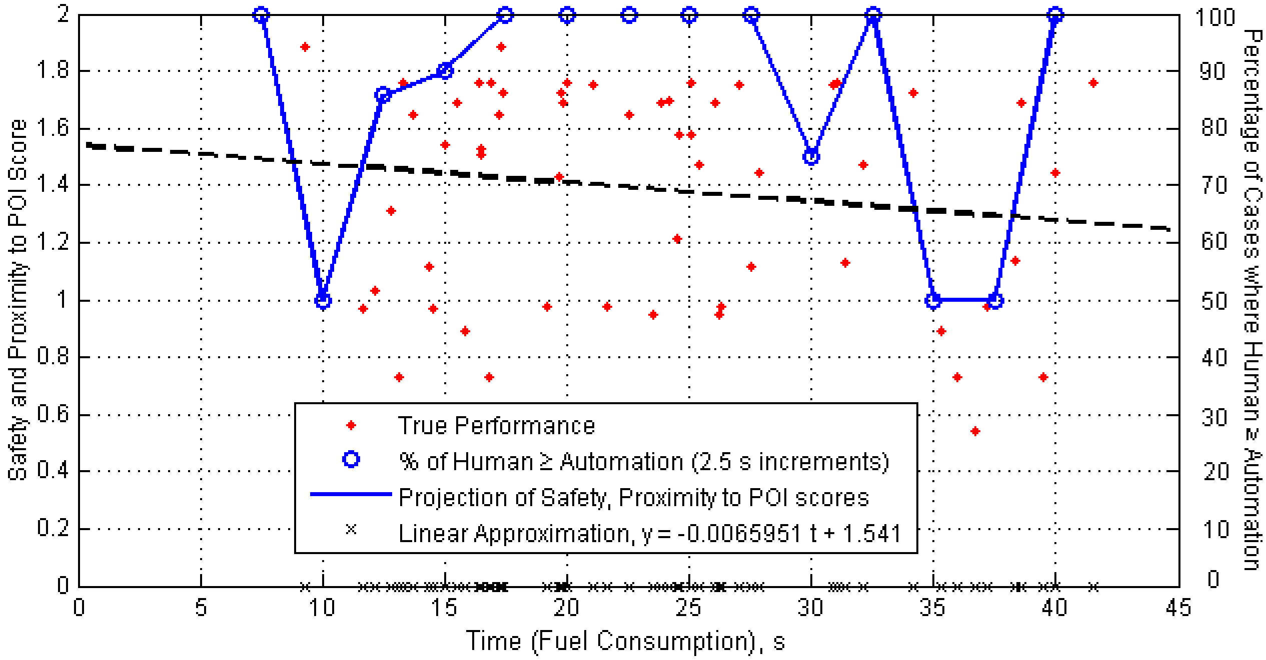


Figure 4 –Human and Automated Site Selection Quality over Time. *The percentage of cases where human site selection is equal to or greater than the automated selection score is determined by selecting all the points within a 2.5 s increment and calculating the percentage with respect to those variables.*

However, this capability comes at a substantial cost to time and consequently fuel. The question remains as to whether the gains in safety and POI proximity justify the additional fuel consumption. Unfortunately, this decision must be made early in the design process, to account for lander fuel tank sizing. The trends illustrated in Fig. 4, while useful in examining the experimental data against a simulated automatic landing system, do not convey the generalized design space. An idealized definition of optimal LPR task performance is retention of human-decision making competency while reducing the required task completion time. Applying this definition to the experimental data illustrates the theoretical human performance in comparison to an automated system. The time to complete can be considered as an independent variable, if the pilots' site selections are assumed to be time-independent and unbiased by time pressure perceptions. This trend is seen in Fig. 5, which compares the relative quality of human control and automated control, allowing either the site selection decision or task completion time to vary for the pilots', the automated system, or both.

The dash-dotted line in Fig. 5 describes the instance where the pilots' site selection in the 56 cases is held constant, but the time to complete the selection for both the automated system and the pilot is matched. This trend represents the situation where an automated system is designed to take over decision authority should the need arise. At zero time to complete, the human performance is equal to or superior to the automated in 79% of the cases. This value indicates that even in the ideal situation, there are still instances in which the pilot has chosen a worse site than the *a priori* selection. However, this percentage of superior or equivalent PPS proportionally increases with the task time, connoting that while the fuel penalty aggravates the PPS and

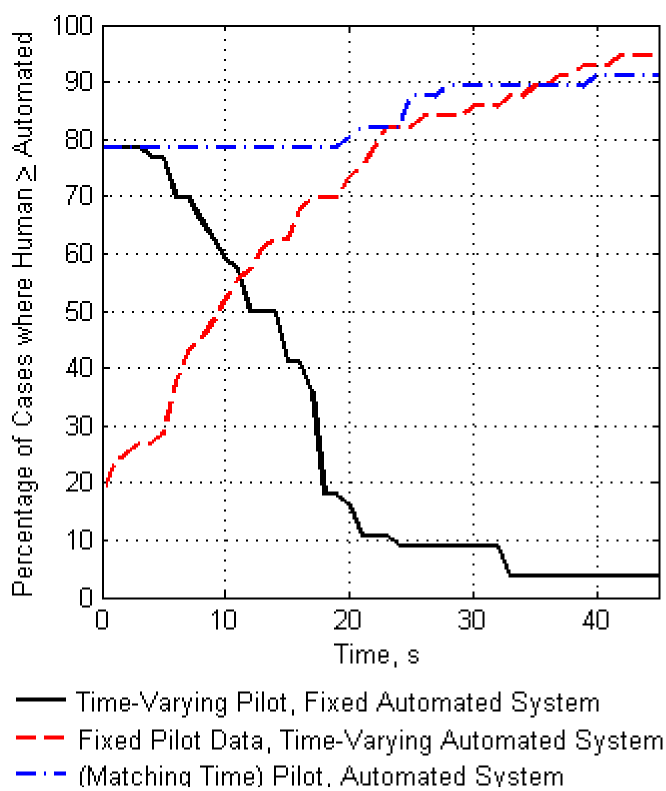


Figure 5 – Comparison of Theoretical, Actual Pilot and Automated System Performance.

APS, the APS score suffers because of the substandard nature of the baseline landing point. Therefore, in order for an automatic landing system to act as a fail-safe option (should the human be unable to make a decision within 20s), the automatic system should possess the capability to select alternative landing sites. Otherwise, an exchange in decision-authority from human to automation after 20s should not be approved, as the quality of the automated decision may not be sufficient to overcome the incurred fuel consumption penalty.

The dashed line in Fig. 5 represents a comparison between the actual pilot data (including true task completion times) and a time-varying automated system. This trend does not physically represent a realistic system, but emphasizes the significance of the fuel consumption penalty. As mentioned previously, at zero time to complete, the human underperforms compared to the automation. However, as the automated system takes more time to complete, perhaps due to sensor and algorithm processing, the APS decreases. At about ten seconds, the experiment PPS begins to score higher relative to the APS. Therefore, if the automated system requires more than ten seconds, the human may be more reliable in choosing an appropriate site.

The last trend observed in Fig. 5 is most representative of a realistic automated system. In this comparison, the automated system requires zero time to complete while the pilots' site selections are held constant while the time to complete for all 56 cases is varied. This solid black line clearly indicates the importance of time/fuel on the overall score. This line, concurrently with the dash-dot line, assumes that all the pilots' can be trained to retain the same level of decision-making integrity in a specified amount of time. At zero time to complete, the human is superior to the automated system in 79% of the cases. The fuel penalty affects the PPS at a generally constant rate until about 12s (the time at which the fastest pilots in this study began to return decisions). At this point, the automated system begins to outperform the human and the likelihood of gaining mission and safety advantages is significantly decreased. The trend demonstrates distinctive curvature due to the coupling of fuel consumption and divert maneuvers to reach sites located on the extremes of the landing area. At 33s, the percentage has dropped to 4%, rendering the human system ineffective relative to the automated system. Therefore, with respect to the reference automated system used in this analysis, the pilots would need to be trained to complete the LPR task in less than 12s in order to provide a decision-making advantage over an automated system.

The observations noted in Fig. 5 are valid only for one specific weighting criterion, where safety, proximity to POI and fuel consumption are equally important. To account for the variability in weighting distribution on the measures of interest, the solid black line in Fig. 5 was re-examined over the full design space of mission criteria. The PPS for the 56 cases was determined using Eq. 1. As illustrated in Fig. 6, one weight (w_i) was incremented by 0.1 while the other two

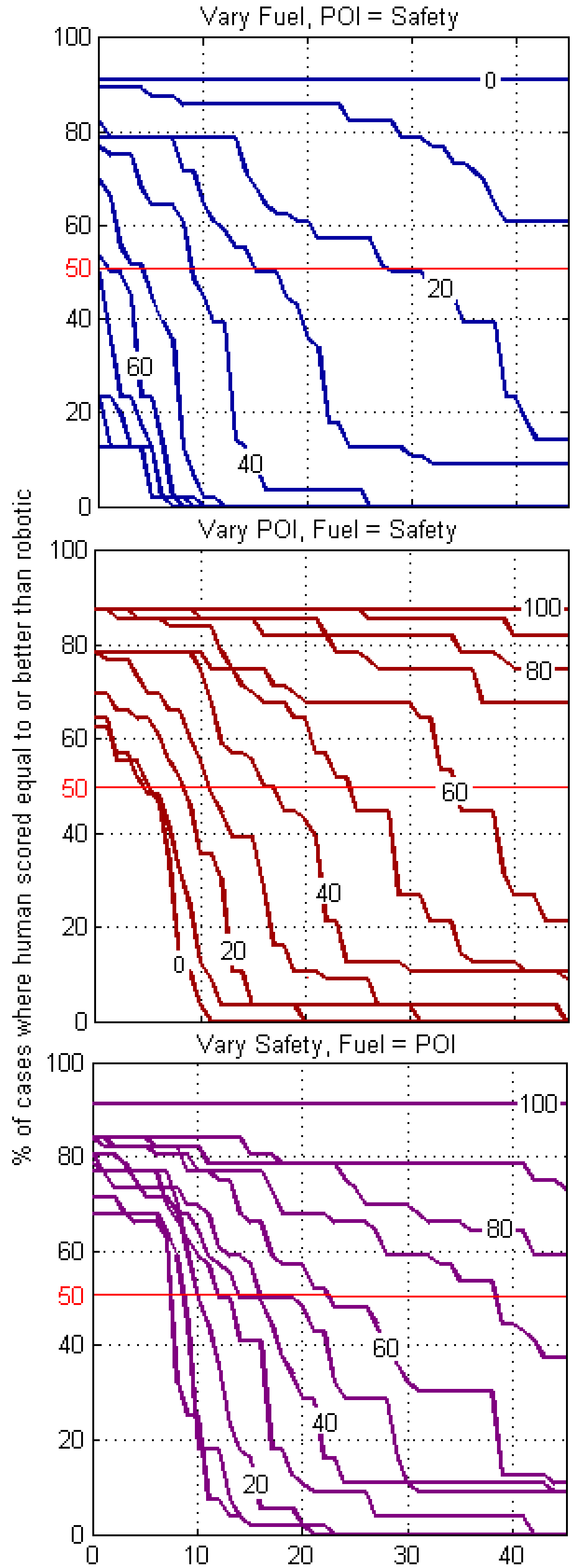


Figure 6 – Comparison of Automated and Human Performance Scores under Variable Definitions of Mission Success.

weights were equivalent and equal to $0.5(1-w_i)$. Similarly to Fig. 5, the pilots are assumed to retain the same level of decision-making capability while performing the LPR at some arbitrary time period and the automated system always selects the baseline point in zero seconds.

Each line in Fig. 6 represents one definition of mission success, i.e., one combination of metric weights. The most distinctive feature of these contours is that no definition exists where the human always outperforms the automated system. The maximum level (91%) of human performance occurs when safety is the only objective of concern. A similar situation occurs when proximity to POI is the primary objective function. At this definition, the human pilot has selected a site closer to the POI in 88% of the cases. The human pilot fares well when both POI proximity and safety are equally critical and fuel consumption is of no importance, selecting better sites than the automated landing system in 91% of all cases.

The same general pattern and relationship with fuel consumption is seen in all three graphs of Fig. 6. As task completion time increases, the fuel penalty grows proportionally until a point where the human pilot cannot perform with the same quality as the automated system. This critical point, at the extremes, occurs at 7s, 11s, and 21s for the fuel consumption, proximity to POI, and safety plots, respectively. Referring to the task completion time distribution in Fig. 4, the contours in Fig. 6 illustrate that an automated system is a more suitable choice for missions where fuel consumption is weighed heavily ($w_f > 0.5$), as the human pilot cannot generally make better decisions than the automated landing system.

In many ways, these contours are analogous to cumulative distribution functions (CDF). If the 56 cases performed in this experiment are representative of typical site selection choices during LPR, then these contours can assist in determining the conditions for human control. For example, if the lunar vehicle was constrained to hold enough fuel for 20s of LPR, then the mission designer can compare human and automated performance for different combinations of weighting distributions. These contours show that if the primary driving factors are safety, proximity to POI, or both, at weighting distributions of at least ≥ 0.7, 0.6, or 0.45 each, respectively, then the human will choose a better landing site than one chosen *a priori* in at least 60% of the cases. While these contours are useful in providing initial estimations to the LPR task, one should note that the modeled human performance may not be accurately representative of actual astronaut behavior and that an automated lander would most likely not operate under the same conditions as a crewed lander. Additionally, the automatic landing system was assumed to have 100% reliability. In the future, true CDF plots produced using trained astronauts and accurate unmanned trajectories should be used.

The analysis thus far has focused on an automated system that was incapable of selecting the best landing site. The experimental results provide sufficient data to analyze the instantiation of an ideal automated system. The scenarios involving the *a priori* sites as the top ranked site were examined, to determine the likelihood of pilots' not recognizing the best site (or an equivalent) and selecting poorer touchdown points. Fifty-seven cases were used in this analysis. Under the equalized weighting distribution used for this experiment, the pilots were able to correctly identify the *a priori* site as the top site in 62.5% of the cases. This percentage of identification shifted with respect to the changes in weighting distribution. Similar to the case of the non-ideal automated system, the human performs best compared to the automation in 85.7% of the cases when the mission singularly emphasizes safety. Additionally, the human is unable to reproduce the same level of performance as the automation due to fuel consumption. This critical point occurs at 5s, 9s, and 10s for the fuel consumption, proximity to POI, and safety plots, respectively.

5. DISCUSSION

The experimental data collected in this study was used to examine several theoretical situations in order to begin quantifying the advantages and disadvantages in human control during LPR given the limited set of human performance information available. Mission designers may find the results of this study useful during conceptual design, for determining the human pilots' role during LPR. From this analysis, it is clear that the human pilot is capable of finding landing sites that are congenial to vehicle safety and mission success. However, given any realistic mission scenario, this landing site decision-making process must occur quickly, otherwise an extensive fuel consumption penalty is invoked.

There are several possible strategies mission designers can employ to reach maximized LPR task performance. First, the reference automated system used in this analysis assumed the system that did not receive any real-time data. The baseline point is formulated on *a priori*, pre-launch mission data. Therefore, the trends presented would shift significantly if a real-time automated decision-making algorithm was used. Implementing an improved algorithm, however, may introduce complexity and additional costs, such a processing time and memory capability.

Second, astronauts could be trained to complete the LPR task in a fixed period of time. This method has been used successfully during the Apollo missions and will most likely be employed on future missions. This study illustrated no significant correlation between time to complete and site selection (with respect to safety and proximity to POI only). Consequently, the possibility exists that through training and personnel selection, better performance could be achieved in shorter time periods. Thus, a more useful solution is to determine the time period necessary for well-trained individuals to make an LPR decision and to train astronauts under those time constraints to determine their likely performance. This study indicates that even relatively untrained individuals can make diagnoses in 12-28s and that substantial fuel induced penalties are likely to become critical as early as 7s, depending on the metric weighting combination.

Lastly, the specific role of the astronaut may need to be adjusted. This experiment scenario was designed such that the pilot would be responsible for evaluating a number of automation-suggested landing sites and making a final decision. The pilot was also told there was no "fail-safe" mode – the absence of a decision would result in a mission abort. A different level of responsibility may result in improved decision-making capability. For example, improved pilot performance may occur if multiple redesignation opportunities are available.

6. Future Work

Although the experiment results shed insight to human performance during the LPR task, additional studies are required to improve the fidelity of existing LPR human-system interaction models. These additional studies should focus on deriving a more explicit relationship between mission and environment inputs and pilot responses, as manifested by interactions with a reference lunar landing system. Employing astronauts in lieu of recreational pilots in future studies would minimize the discrepancy and ambiguity regarding a standard definition of representative astronaut behavior. However, running human-in-the-loop experiments with astronauts is impractical for performing comparative trade studies between humans and automated flight systems. A probabilistic computational human performance model (CHPM) is needed, to examine the full discourse of astronaut behavior in a wide array of landing scenarios within the time constraints. A significant experimental effort is necessary to validate such a probabilistic model, but to rely strictly on human subject testing would prove to be an expensive and time-consuming task. Furthermore, a CHPM may be adapted to examine lunar landing operation during underperforming or impaired pilot performance, scenarios that are difficult to emulate under laboratory conditions. The CHPM is not intended to replace astronauts and cannot be guaranteed to emulate their behavior in every scenario, but should provide an accurate approximation on the performance of crewed lunar landing and provide insight as to when the use of more in-depth human-in-the-loop study is merited. These approximations should be of use to mission designers during systems architecture studies.

7. Conclusion

The landing point redesignation task permits onboard crew to evaluate and select alternative sites prior to terminal descent and touchdown. During this opportunity, the crew must balance the safety of the vehicle and the goals of the mission without violating fuel constraints. Given the complexity of this task, it is likely that the responsibilities during this task will be shared by an automated system. However, quantifiable methods are needed to appropriately partition work to achieve robust or optimal combinations of human-system interaction. One of these methods consists of comparing human and automated performance with respect to task completion time. Based on the results of this experiment, humans tend to select better landing sites than the reference automated system when safety and proximity to points of interest are the most critical criteria. However, the decision-making time required for humans incurs significant fuel consumption costs. Thus, in landing scenarios when reserving fuel is of greater priority, mission designers may opt to limit human control during landing site selection. Human pilots were also able to match the performance of a perfect automated system for more than half of the examined cases, but were also prone to diverting to worse landing sites. Adjustments to astronaut training or improvements to onboard decision-making aids would enhance the synthesized site-selection performance.

The experiment results indicate a need to improve human-system interaction modeling with increased correlation between mission scenario and human performance. A probabilistic computational human performance model would be more conducive to generating the quantity of data necessary to observe quantified approximations on crewed lunar landing performance. The observations gleaned from this analysis lay the foundation for future investigations into the specific region of optimal human control. Designing the human role to account for the prime behavior should reduce the risk for violating system capabilities while allowing astronaut input.

References

[1] L. Bainbridge. Ironies of Automation. Automatica, Vol. 19, 1983. pp. 775-779.

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

[3] Z. Chua, K. Feigh. Investigation of Pilot Behavior during Landing Point Redesignation. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans. Submitted September 1, 2009. In Review.

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

[5] L. J. Kessler and L. M. Forest, "Human-interactive autonomous flight manager for precision lunar landing," Presentation, Fall 2006, at AAAI Fall Symposium on Spacecraft Autonomy.

[6] L. Forest, L. Kessler, and M. Homer, "Design of a human-interactive autonomous flight manager (AFM) for crewed lunar landing," in AIAA Infotech@Aerospace, 2007.

[7] J. Davis, S. Striepe, R. Maddock, G. Hines, S. P. II, B. Cohanim, T. Fill, M. Johnson, R. Bishop, K. DeMars, R. Sostaric, and A. Johnson, "Advances in POST2 end-to-end descent and landing simulation for the ALHAT project," in AIAA/AAS Astrodynamics Specialist Conference & Exhibit, no. AIAA 2008-6938, 2008.

[8] F. Bennett. "Apollo Lunar Descent and Ascent Trajectories." Technical Report. NASA, Houston TX. March 1970. Also presented at AIAA 8th Aerospace Sciences Meeting, January 1970.

[9] Z. K. Chua, L. M. Major, and K. M. Feigh, "Modeling cockpit interface usage during lunar landing redesignation," in International Symposium of Aviation Psychology, 2009.

[10] Z. Chua, R. Braun, K. Feigh. Analysis of Human-System Interaction for Landing Point Redesignation. Technical Report. May 2009. Available online. http://hdl.handle.net/1853/29921

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

[12] L. Forest, B. Cohanim, and T. Brady, "Human interactive landing point redesignation for lunar landing," in Proceedings of the IEEE Aerospace Conference, Big Sky, MN, March 2008.

[13] R. Monsen. (2007, October) Eaglelander3d. [Online]. Available: http://eaglelander3d.com/

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

[15] Manned Spacecraft Center. "Apollo 11 flight plan," National Aeronautics and Space Administration, Houston, TX, Flight Plan. AS-506/CSM-107/LM-5, July 1969.

[16] R. Hirsh, NASA Johnson Space Center. September 28, 2009. Personal Correspondence.

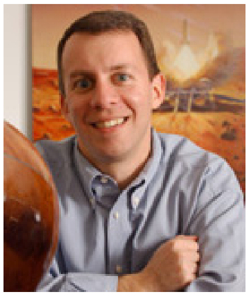
ACKNOWLEDGEMENTS

The authors acknowledge the entire ALHAT project team consisting of members from NASA's Langley Research Center, Jet Propulsion Laboratory, Johnson Space Center, C.S. Draper Laboratory, and the Applied Physics Laboratory. The authors also thank the members of the Space Systems Design Laboratory and Cognitive Engineering Center for their continued support and assistance during this study.

BIOGRAPHY

Zarrin K. Chua *is a student member of both IEEE and AIAA. She holds a B.S. degree from the Virginia Tech and a M.S. degree from the Georgia Institute of Technology, both in aerospace engineering. She is currently a third-year Graduate Research Assistant in the Space Systems Design Laboratory and the Cognitive Engineering Center at the Georgia Institute of Technology in Atlanta, Georgia. She has previously worked as a Student Intern at the NASA Johnson Space Center, the NASA Langley Research Center, and the Charles Stark Draper Laboratory. Her research interests include entry, descent and landing and human performance modeling. Ms. Chua is currently a NASA Graduate Student Researchers Program (GSRP) fellow. Ms. Chua is also a student member of the Human Factors and Ergonomics Society.*

Karen M. Feigh *is an AIAA member. She holds a B.S. in Aerospace Engineering from the Georgia Institute of Technology, an MPhil in Aeronautics from Cranfield University, UK, and a Ph.D. in Industrial and Systems Engineering from Georgia Institute of Technology. Dr. Feigh is an assistant professor at the Georgia Institute of Technology's School of Aerospace Engineering. She is currently the Acting Director of Georgia Tech's Cognitive Engineering Center. She has previously worked on fast-time air traffic simulation, conducted ethnographic studies of airline and fractional ownership operation control centers, and designed expert systems for air traffic control towers. Dr. Feigh's research interests include the design of cognitive work support systems for individuals and teams in dynamic socio-technical settings, including airline operations, air transportation systems, UAV and MAV ground control stations, mission control centers, and command and control centers. Her awards include the Marshall scholarship and the AIAA Orville and Wilbur Wright Graduate Award. Dr. Feigh is also a member of the Human Factors and Ergonomics Society.*



Robert D. Braun *is the David and Andrew Lewis Associate Professor of Space Technology in the Daniel Guggenheim School of Aerospace Engineering at the Georgia Institute of Technology. As Director of Georgia Tech's Space Systems Design Laboratory, he leads a research program focused on the design of advanced flight systems and technologies for planetary exploration. He is responsible for undergraduate and graduate level instruction in the areas of space systems design, astrodynamics, and planetary entry. Prior to coming to Georgia Tech, he served on the technical staff of the NASA Langley Research Center for sixteen years, where he contributed to the design, development, test, and operation of several robotic space flight systems. He has worked extensively in the areas of entry system design, planetary atmospheric flight, and mission architecture development. Dr. Braun is an AIAA Fellow and the principal author or co-author of over 150 technical publications in the fields of planetary exploration, atmospheric entry, multidisciplinary design optimization, and systems engineering. He has a B.S. in Aerospace Engineering from Penn State University, a M.S. in Astronautics from George Washington University, and a Ph.D. in Aeronautics and Astronautics from Stanford University.*