Level of Automation and Failure Frequency Effects on Simulated Lunar Lander Performance

Jessica J. Marquez
NASA Ames Research Center
Moffett Field, CA 94035
Jessica.J.Marquez@nasa.gov

Margarita Ramirez
San Jose State University
NASA Ames Research Center
Moffett Field, CA 94035
Margarita.Taula@nasa.gov

Abstract— A human-in-the-loop experiment was conducted at the NASA Ames Research Center Vertical Motion Simulator, where instrument-rated pilots completed a simulated terminal descent phase of a lunar landing. Ten pilots participated in a 2 x 2 mixed design experiment, with level of automation as the within-subjects factor and failure frequency as the between-subjects factor. The two evaluated levels of automation were high (fully automated landing) and low (manual controlled landing). During test trials, participants were exposed to either a high number of failures (75% failure frequency) or low number of failures (25% failure frequency). In order to investigate the pilots' sensitivity to changes in levels of automation and failure frequency, the dependent measure selected for this experiment was accuracy of failure diagnosis, from which D Prime and Decision Criterion were derived.

For each of the dependent measures, no significant difference was found for level of automation and no significant interaction was detected between level of automation and failure frequency. A significant effect was identified for failure frequency suggesting failure frequency has a significant effect on pilots' sensitivity to failure detection and diagnosis. Participants were more likely to correctly identify and diagnose failures if they experienced the higher levels of failures, regardless of level of automation.

TABLE OF CONTENTS

1. INTRODUCTION

Future space exploration missions need to better integrate humans and automation, particularly during spacecraft vehicle control. While advances in automation progress, there is limited research on how best to integrate the pilot with these new aerospace systems. This research (a continuation of [1] and [2]) investigates the lunar exploration mission, simulating landings, a time-critical phase of flight.

Flying a spacecraft is a complex task that involves comprehensive understanding of the instruments and how they affect the spacecraft, constant situation awareness of past states and the current state in order to predict possible future states, evaluation of risks and benefits, and decision-making. It is important that pilots are able to operate under periods of high workload in the event of a malfunction. In many instances, such as entry, descent, landing, and docking, flying may be coupled with time-critical situations where the choices made in a matter of seconds can determine mission success or failure.

Even during Apollo 11, astronauts encountered human-automation integration issues. Some of the complications that occurred include computer information overload resulting in alarms sounding and software restarts, instability of the throttle control algorithm causing the descent engine to oscillate uncontrollably, and even loss of communication [3]. This illustrates the significance of training pilots to understand how the automated system works.

High cognitive demand is placed on pilots during spacecraft vehicle control as they must detect, diagnose, and recover from system failures. For lunar terminal descent, the challenge is exacerbated due to the limited time available to recover from such failures. Additionally, the amount of automation provided to the pilot in turn impacts these demands. The degree of trust an operator has in the automation and the automation's failure frequency is associated with the effective use of these systems. If the automation fails often, the operator will be reluctant to make use of it, but if the automation is impeccable, the operator will rely heavily on it [4]. Failure frequency is only one of many factors influencing operator use of automation.

It is important to understand the conditions that influence a pilot's confidence in the automation and a pilot's ability to detect failures to mitigate potential complications that may occur as a result, such as decreased situational awareness and overall mission abortion. Hence, our research goals were to evaluate pilot failure detection and diagnosis, during flight and landing on the lunar surface, as they adapted to spacecraft failures, failure frequency, and different spacecraft levels of automation.

2. HUMAN-AUTOMATION INTERACTION

The term "automation" has been refined and redefined over the last 60 or so years. Sheridan defines automation as "...any mechanical or electronic replacement of labor, where labor is taken to mean either physical labor or mental labor" [5]. Today the term is used across domains such as space, aviation, automotive, medical, and home and entertainment; its meaning has expanded to include 1) artificial sensory mechanization and integration, 2) computer processing, 3) mechanical activity, and 4) "information action" [6]. In the current experiment we apply Sheridan's definition of automation and split the labor between high or low automation, which are later explained in more detail.

Issues with human-automation interaction have become more salient as automation became more complex without correspondingly investigating the effects this had on the person's ability to control vehicles. For instance, in 1997, while on approach for landing, a commercial aircraft crashed 19-miles southwest of Detroit Metropolitan Wayne County Airport [7]. The crew was operating on autopilot and did not understand the autopilot state, so their adjustments to the autopilot worsened the controllability of the airplane. According to the Flight Deck Automation Issues Database [8], reduced situation awareness and pilots being out-of-the-loop were contributing factors in this incident. Perhaps if the pilots had been flying in a manual mode, they would have noticed the change in dynamics (of icing effects) and would have been able to intervene prior to any loss of control. This incident is indicative of the significance that the level of automation places on an operator's ability to detect an anomaly and correct the issue in a timely manner.

3. FAILURE DETECTION

Failure detection and diagnosis is key in space and other high risk and/or time sensitive settings, such as those encountered in search and rescue operations, aviation, military operations, medicine, and nuclear power plant management. Timely failure detection provides operators with potentially more time to diagnose and intervene before the situation escalates from manageable to fatal. For instance, in aviation, pilots must be able to quickly detect an anomaly, such as with the Flight Management System (FMS), and take the appropriate steps to rectify the issue. Pilots must detect a target stimulus among background noise, in this case, an anomaly within the large, complex set of system information.

Performance of failure detection tasks in human-automation interactions has been used to measure the effects of automation on the operator, such as complacency [9, 10, 11] and reliance. Complacency can occur when an operator's role goes from manually being in control to simply overseeing a highly reliable system [9]. When an operator becomes complacent, dependency on the automation to alert of system inconsistencies (whether accurate or faulty) occurs, thus allowing for vigilance to decrease. Reliance refers to the operator's dependence or trust when the automation is functioning without alarms, indicating there are no problems. Reliant operators have the mental resources available to attend to other tasks because they are dependent on the automation to alert them when a problem arises. Previous studies have suggested that reliance and compliance are independent constructs. In a study to investigate the relationship between automation reliance and compliance, Dixon, Wickens, and McCarley [12] found that operator performance is affected by automation false alarms as a result of reduced operator reliance. The frequency that operators encounter failures may also affect operator reliance and compliance.

4. AUTOMATION INTEGRATION

With increased automation of systems, the responsibility of the human operator has evolved to meet these changing roles. The shift in function allocation, though beneficial, has not come without obstacles. For example, a pilot monitoring a dynamic system (e.g. autopilot) must detect subtle changes that may not be as easily detected as if he were actively in control [13].

Supervisory Control of Automation

As evidenced in prior research, the human role as a supervisory controller or monitor can account for degradation in situational awareness and skill, system over-reliance or complacency, mode-related errors, and ultimately mission failure [9, 11, 14, 15, 16]. These are consequences of changing the human role from active operator to monitor. In a field study by Wiener [17] to evaluate new glass cockpit technology, pilots' task was to monitor complex or dynamic systems. In situations where the automation reverted back to the human for decision-making, the pilots reported being caught off guard with regards to the status of the system, a phenomenon referred to as automation surprise [18]. Additionally, pilots reported feeling out-of-the-loop and described loss of situational awareness [17].

Levels of Automation

Levels of automation indicate the degree to which the human operator or the automation has control or authorization over specific tasks. Various levels of automation taxonomies have been developed to examine the most effective approach to achieve goals relative to the domain at hand [19, 20, 21, 22]. Transitioning between levels of automation is an area of on-going research.

A common condition that manifests when operators become complacent is referred to as out-of-the-loop unfamiliarity (OOTLUF) [23]. OOTLUF refers to the phenomenon of operators working in conjunction with automation and who do not perform well when having to take over manual control in the event of an automation malfunction. This occurs because the operator becomes accustomed to passively monitoring the system, so when their role instantaneously changes, their skills and situational awareness may have deteriorated.

In a study to analyze OOTLUF and information sampling strategies, Lorenz, Di Nocera, Rötger, and Parasuraman [24] studied the performance of 24 participants on three tasks: 1) monitoring system states, 2) alarm reaction time, and 3) memory task. All participants experienced each of the three tasks, while the level of automation condition was a between-subjects factor. Results indicated that automation improved operator performance and reduced operator workload. There was a noticeable decrease in performance in the high-level automation condition when the automation failed. Additionally, information sampling strategies suggested that level of automation helped to keep task emphasis from varying so much, thus retaining situation awareness. Overall, the high-level of automation condition left operators with more cognitive resources to oversee the system.

Situation Awareness

Situational awareness (SA) is the perception, comprehension, and projection of elements in the operational environment [25]. Maintaining SA contributes to piloting task successful performance. Lack of SA, in turn, contributes to loss of vehicle and life, for example [26].

Research has shown that higher levels of automation are associated with out-of-the-loop syndrome, the consequence of complacency and degradation in skill and situation awareness, resulting from prolonged supervisory control of automation [15]. Furthermore, the greater trust an operator places on automation, the less likely the operator is to attend to the system and detect an anomaly. In a study of detection of changes in system dynamics, Kessel and Wickens [13] compared detection performance between operators controlling and monitoring system dynamics. Results from this study indicate that detection is slower when an operator is in an automatic mode, in comparison to a manual control mode. Additionally, there is a positive-transfer from the

manual mode to the automatic mode, which implies that it is best to train operators in a manual level of automation because they will be more perceptive of system changes once they transfer to a more automatic level of automation. Parasuraman et. al. [11] suggested that having operators intermittently take over control of automated systems improves failure detection. Further, imbedding artificial failures at arbitrary times, during non-critical flight periods, may help decrease operator complacency.

5. METHOD

Apparatus

To evaluate failure frequency throughout different levels of automation and how failure frequency and level of automation affect operators' flying performance in terms of safety and failure detection and diagnosis, an experiment was conducted in the NASA Ames Research Center Vertical Motion Simulator (VMS). The experiment set-up followed a similar setup to that of Kaderka, 2013 [2].

The VMS consists of an interchangeable cab that can produce high fidelity, real-time piloted simulations with appropriate motion cueing, resulting in realistic sensory feedback relative to a vehicle's flight characteristics. The motion simulator has a range of motion of up to 60 feet vertically, 40 feet horizontally, and 8 feet laterally. The interior cab is modified to replicate aerospace vehicles' flight controls and displays and out-the-window views [27].

The Interchangeable Cab (ICAB) was customized in the interior by installing flight controls, flight instruments, and aircraft seats to emulate Apollo lander spacecraft. The lunar ICAB was placed on the six degrees-of-freedom motion-based simulator.

The cockpit displays that participants used to navigate the lunar lander included the Horizontal Situation Display (HSD), Primary Flight Display (PFD), and Landing Area Display (Figure 1).

Figure 1 – Cockpit displays layout

3
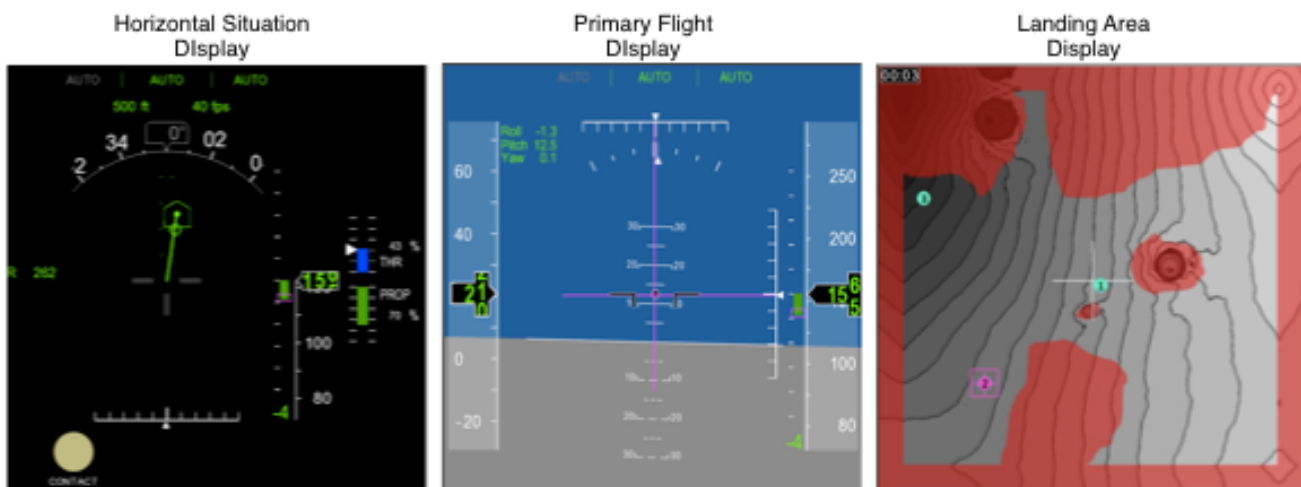
The HSD provided a pictorial view of the vehicle's position relative to navigation points, a visual depiction of navigation and control parameters, such as translational velocity, distance to target, throttle and propellant usage, and rate of descent tape. The PFD provided information about vehicle states such as attitude and horizontal and vertical velocity and guidance through the dual cue flight director. The Landing Area Display provided information about hazards, recommended landing aimpoints, and the amount of fuel remaining.

The Translation Hand Controller (THC) was used to control rate of descent by pulling up or pushing down. The Rotation Hand Controller (RHC) was used to toggle through landing aimpoints during landing aimpoint selection (use of trigger), change modes after landing aimpoint selection (use of slider button), detect and diagnose failures (use of trigger), control pitch and roll (left/right movement of the stick), and abort trial (top button).

Independent Variables

Two independent variables were investigated in this experiment: level of automation and failure frequency. The level of automation factor was within-subjects and included: high level of automation and low level of automation. Prior research indicates there is a significant difference in the operator's workload from the high level of automation condition to the low level of automation condition, with highest workload experienced when the operator is responsible for controlling pitch, roll, and rate of descent (low level of automation) [1, 2]. Since participants reported higher workload in the low level of automation condition [1, 2], we expected that accuracy of failure detection would be greater in the condition of lower workload (high level of automation), where pilots may have spare capacity to detect and diagnose failures.

The second independent variable was failure frequency. Participants experienced either a high failure rate (i.e., the probability of a failure occurring throughout the set of experimental trials was 75%) or low failure rate (i.e., the probability of a failure occurring throughout the set of experimental trials was 25%). This two-level factor was a between-subjects variable; during the test session, participants only experienced a 75% or 25% failure rate, which was randomly assigned.

In this experiment, the failures were system failures rather than automation failures. Automation reliability refers to the "number of correct operations done by a computer out of the total number of operations" [28]. For the purpose of this experiment, automation reliability was held constant at 100%. Accordingly, if the automation indicated a failure, then in fact, a failure was occurring; likewise, if the automation showed no manifestation of an anomaly, then no failure had occurred.

Types of Failures

Participants' task was to detect and diagnose system failures. Pilots were presented with two types of failures,

equally distributed regardless of the failure frequency. If a failure was detected, participants had to determine whether the anomaly they experienced was a result of a thruster or radar failure.

A thruster (4C) failure occurred when a thruster "failed on" upon firing (Figure 2). The thruster failure was noticeable through motion cues and abrupt movements in the attitude indicator. In the low automation case, the failure produced a pitch down, roll left response, in which the participant had to manually input commands to null guidance errors in the PFD. In the high automation case, the thruster failure was recognized by the control system through errors in pitch and roll. When the errors exceeded the ±5 degree dead band, the control system attempted to null the errors in pitch and roll present in the guidance needles.
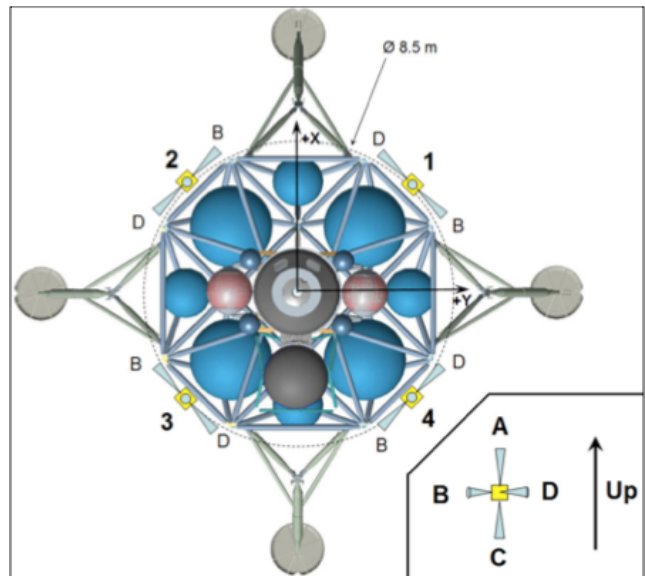

Figure 2 – Overhead view of Reaction Control System jets, with the inset depicting the "outside" view of the thruster pod [29]

A radar failure, or radar malfunction, resulted in additional noise in the altitude tape (in the PFD), rate of descent tape (in the PFD), and horizontal velocity vector (in both the HSD & PFD). The additional altitude noise was integrated by the vehicle's automation to compute the guidance, despite the vehicle's attitude truly remaining constant. In the low automation case, participants had to manually input commands to null guidance errors. In the high automation case, the control system allowed for the radar failure to exhibit itself in the integrated guidance before correcting itself (once reaching the ±5 degree dead band). The radar failure was noticeable through the oscillating errors in the guidance needles, attitude indicator and increased noise in the rate of descent tape and horizontal velocity vector.

Dependent Variables

In order to investigate if level of automation had an effect on failure detection based on failure frequency, correct detection and diagnosis of failures was recorded. For each trial, a "hit" was defined by correct detection and correct

diagnosis when a failure was present. A false alarm was defined as a detection and diagnosis of a failure when a failure was not present. Signal detection statistics were derived from the raw data and used as dependent variables.

D prime (d') is a signal detection statistic that measures a participant's sensitivity to a target stimulus from background stimuli. A larger d' value implies the signal can be more easily detected. An estimate of d' was calculated for each participant by measuring the distribution of hits and false alarms using the following formula: $d' = Z$(hit rate) – Z(false alarm rate).

Decision criterion is a signal detection statistic that measures a participant's willingness, or minimum level of internal certainty, in deciding that a signal is present amongst background stimuli. A larger criterion value implies the participant required stronger evidence before determining that a signal was present. A criterion estimate was calculated for each participant using the false-alarm rate and noise distribution.

Participants

Ten instrument-rated pilots completed a simulated terminal descent phase to the moon, one female and nine males. Participants' age was between 22 and 38 years. All participants were required to have 20/20 vision or corrected-to 20/20 vision. The participant population included test, private, recreational, and commercial pilots, and certified flight instructors; several participants held multiple certifications. Participants were recruited from local aviation companies and all were compensated. NASA Ames Research Center and San Jose State University Institutional Review Boards approved the protocol.

Experimental Design

Participants were randomly assigned to the 25% or 75% failure frequency group, with half of the participants (*n*=5) in each group. The total number of trials varied from 66 to 78 because some participants required more practice trials during training. In the test session, all participants experienced 32 experimental trials. The two failures (thruster and radar) were evenly distributed among each participant's failure trials (i.e. 25% failure frequency = 8 failure trials = 4 thruster failures, 4 radar failures; 75% failure frequency = 24 failure trials = 12 thruster failures, 12 radar failures). Failures randomly occurred 45 to 60 seconds after the trial onset. All participants' test session trials were evenly split to allow them to experience 16 of the trials in the high automation condition and 16 in the low automation condition. While maintaining this distribution, the experimental trials were randomized within the test session.

Procedure

The experiment was conducted in four parts: 1) questionnaire and experiment familiarization briefing, 2) training session where participants were familiarized with the experimental tasks and protocol, 3) test session where participants experienced experimental conditions and their

performance data was recorded for analysis, and 4) post experimental interview.

At the start of each session the participant was provided an informed consent form. The participant was then asked to answer questions about his/her piloting experience, motion sickness susceptibility, and standard demographic information. Following the questionnaire, participants were acquainted with the experiment details and their mission objectives, during which participants were encouraged to ask questions to ensure their understanding of their role in the experiment.

After a VMS safety briefing, the experimental training session began, in which the participant learned to operate the lunar lander simulator. During the training session, trials initially did not include motion nor failures to allow participants to get acquainted with the controls and timing of events. Failures were then incorporated into the training trials and participants were taught how to recognize them. During training, all participants were exposed to a 50% failure rate, independent of their test condition (failure frequency group). This setup was important so as to not set up an expectation for a particular failure frequency. Throughout all the training trials, participants were provided feedback with respect to their piloting method, callouts of vehicle states (i.e. altitude, fuel, and landmarks), and failure detection and diagnosis. All participants' training session, regardless of failure frequency condition, were conducted in the same manner. The number of sufficient training trials was determined from previous experiments [2].

Each trial began with landing aimpoint selection, in which participants' task was to select the safest landing aimpoint from several candidate points, which were overlaid on a terrain map (Figure 1). Participants were instructed to select the safest landing aimpoint (i.e., furthest from a hazard, or any red area). During this phase, participants were not required to input control inputs, irrespective of the level of automation condition. After the landing aimpoint selection phase, the participant flipped the transition switch to change the Landing Area Display from a stationary map to a moving map and transition to the experimental control mode, which changed to low automation or remained in high automation. In the low automation condition, the pilot was responsible for commanding pitch, roll, and rate of descent during the final descent phase, in addition to detecting and diagnosing systems failures and attending to the callouts. In the high automation condition, flying continued to be performed by the autopilot during the final descent phase, leaving the pilot to detect and diagnose system failures and attend to the callouts.

The callouts were incorporated to ensure participants were cognizant of their situational awareness as well as to maintain consistent workload across similar experiments [2]. Given that verbal callouts are routine for pilots to perform, the verbal callout method of assessing situation awareness has high face validity [2, 30]. Throughout both control modes, participants' remained vigilant to the task of

failure detection and diagnosis, while attending to the callouts.

During the test session, all participants experienced 32 experimental trials, each high automation trial lasting 80 seconds and low automation trials lasting 95 seconds. Failure timing randomly occurred between 45 to 60 seconds after the trial onset.

6. RESULTS

The research goals were to evaluate pilots' failure detection performance during flying and landing on the lunar surface as they adapted to spacecraft failures, failure frequency, and different spacecraft levels of automation. The purpose of this experiment was to measure failure detection performance of participants as they flew the lunar lander simulator under two levels of automation (high automation and low automation) and two levels of failure frequency (75% failure rate and 25% failure rate). Signal detection theory (SDT) was the underlying method used to evaluate the ease or difficulty of detecting a failure from among background noise (d') and participants' decision criteria in determining whether a failure had occurred (criterion). The rationale in applying SDT was to quantify participants' ability to detect failures, that is, measure how they make decisions under ambiguous conditions. A standard correction was first calculated for hit rates of 1.0 and false alarm rates of 0 [31]. Two-way mixed measures ANOVAs were performed to test for mean differences between groups, where the significance level adhered to was $\alpha = 0.05$.

Hits and False Alarms

For each trial, a hit was defined as correct detection and correct diagnosis when a failure was present. A false alarm was defined as a detection and diagnosis of a failure when a failure was not present. In assessing the overall hit rate, it was observed that participants achieved a mean hit rate of 92.5% in the high failure frequency condition and a mean hit rate of 87.5% in the low failure frequency condition (Table 1). In assessing the overall false alarm rate, it was observed that participants achieved a mean false alarm rate of 14% in the high failure frequency condition and a mean false alarm rate of 19% in the low failure frequency condition (Table 2).

Table 1. Hit Rate Means by Condition

Condition		Failure Frequency	
		25%	*75%*
Automation Level	*High*	0.83	0.93
	Low	0.78	0.92

Table 2. False Alarm Rate Means by Condition

Condition		Failure Frequency	
		25%	*75%*
Automation Level	*High*	0.12	0.13
	Low	0.26	0.15

D Prime

Results showed no significant difference for level of automation $F(1, 8) = 1.635$, $p = 0.237$; no significant interaction between conditions, $F(1, 8) = 0.590$, $p = 0.465$; but here was a significant main effect for failure frequency, $F(1, 8) = 5.457$, $p = 0.048$. Overall, the participants who experienced the 75% failure frequency condition had an easier time detecting and diagnosing failures than did participants who experienced the 25% failure frequency condition (Figure 3).
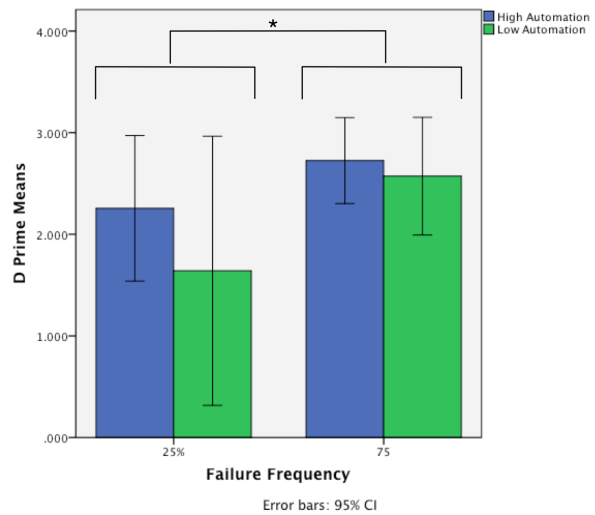

Figure 3 – D Prime mean differences between level of automation and failure frequency (* indicates significance)

Decision Criterion

Decision criterion was used to identify any difference in participants' judgment regarding the presence of a failure. Results showed no significant difference for level of automation $F(1, 8) = 1.495$, $p = 0.256$; no significant difference for failure frequency, $F(1,8) = 0.042$, $p = 0.842$; and no significant interaction between level of automation and failure frequency, $F(1, 8) = 0.662$, $p = 0.440$. The lack of significance for any of the experimental conditions indicates that the participants had relatively equal willingness to detect a failure, regardless of failure frequency and level of automation (Figure 4).
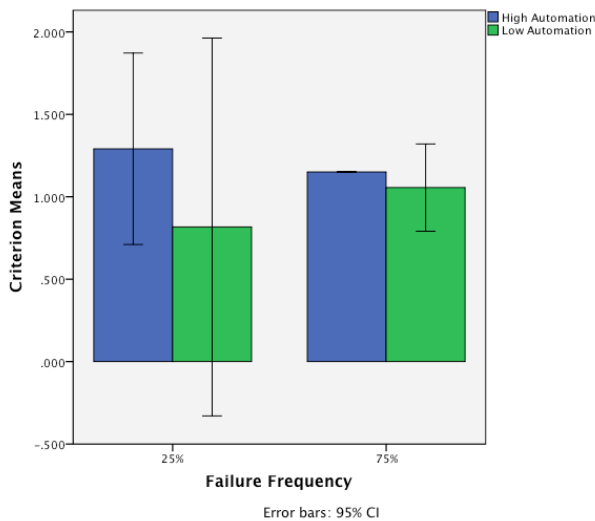
Figure 4 – Criterion mean differences between level of automation and failure frequency

Post-Experimental Questions

After each experimental session, participants were asked questions regarding their experience. Two participants felt they did not have enough training learning to fly the vehicle and only one participant felt he did not have enough training in identifying failures. These participants' data was not removed from the analyses because we felt their flying performance and failure detection performance was satisfactory. Despite the unique flight control task, the majority of participants felt comfortable flying the lunar lander and identifying each of the two failures (radar malfunction and thruster failure).

Participants' responses varied in the post-experimental questions, though there were some responses that did stand out. Across failure frequency groups, a few of the participants commented on their perception of "ease" of failure detection as a function of level of automation. For instance, one participant commented: "It was harder to recognize a failure if you weren't on guidance. It was easier to detect failures on autopilot (high automation condition)." However, experimental results did not support these perceptions.

7. CONCLUSION

The goal of this research was to examine the interaction between level of automation and failure frequency on failure detection and diagnosis. The results from this experiment indicate that failure frequency affects participants' ability to detect failures. Failure frequency could be included as a factor while training in complex, human-automation systems. Training on system failure detection allows operators to better learn and understand the system's strengths and weaknesses [32, 33]. These results can be evaluated in the context of simulated spacecraft training, suggesting that while pilots can effectively learn to detect and diagnose spacecraft failures when exposed to high failure frequencies, that assessment of failure detection and

diagnosis should be completed both at high and low failure frequencies. This type of training may help people better understand automation and gain the appropriate level of confidence, so as to not over- or under-rely on automation.

These results also validate on-going research on failure detection and diagnosis of human supervisory controlled spacecraft. Investigating human performance can be consistently achieved by testing one, consistent failure frequency, preferably one that is high so as to be able to evaluate other effects on failure detection and diagnosis.

8. ACKNOWLEDGEMENTS

REFERENCES

[1] C. J. Hainley Jr., K.R. Duda, C.M. Oman, & A. Natapoff (2013). "Pilot performance, workload, and situation awareness during lunar landing mode transitions." Journal of Spacecraft and Rockets, 50, 793-801. doi: 10.2514/1.A32267

[2] J. Kaderka (2013). "Pilot detection of system failures: experiments and model development." Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

[3] D. Eyles (2004). "Tales from the lunar module guidance computer." *Presented to the 27th Annual Guidance and Control Conference of the American Astronautical Society,* in Breckenridge, Colorado. Retrieved from http://www.doneyles.com/LM/Tales.html

[4] R. Parasuraman (1997). "Humans and automation: use misuse, disuse, abuse." *Human Factors*, 39, 230- 253. Retrieved September 4, 2012, from http://archlab.gmu.edu/people/rparasur/Documents/Paras RileyHF97.pdf

[5] T. B. Sheridan (2002). "What is automation?" *In Humans and Automation: System Design and Research Issues*. John Wiley & Sons, Inc., Santa Monica, CA, 9.

[6] T. B. Sheridan & R. Parasuraman (2005). "Human-automation interaction." *Reviews of Human Factors and Ergonomics 2005*, 1, 89-129. doi: 1518/15575405783703082

[7] National Transportation Safety Board (1998). *In-Flight Icing Encounter and Uncontrolled Collision with Terrain, COMAIR Flight 3272, Embraer EMB-120RT, N265CA, Monroe, Michigan, January 9, 1997*. Aircraft Accident Report NTSB/AAR-98/04. Washington, DC: National Transportation Safety Board.

[8] "Evidence." (2007). Retrieved October 25, 2012, from http://www.flightdeckautomation.com/fdai.aspx

[9] N. Bagheri & G. A. Jamieson (2004). "Considering subjective trust and monitoring behavior in assessing automation-induced 'complacency'." *In Proceedings of the Human Performance, Situation Awareness and Automation Conference*, Marietta, GA.

[10] S. R. Dixon & C. D. Wickens (2006). "Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload." *Human Factors*, 48, 474-486.

[11] R. Parasuraman, R. Malloy, & I. L. Singh (1993). "Performance consequences of automation induced 'complacency'." *The International Journal of Aviation Psychology*, 3, 1-23.

[12] S. R. Dixon, C. D. Wickens, & J. S. McCarley (2007). "On the independence of compliance and reliance: are automation false alarms worse than misses?" Human Factors, 49, 564-572. doi: 10.1518/001872007X215656

[13] C. J. Kessel & C. D. Wickens (1982). "The transfer of failure-detection skills between monitoring and controlling dynamic systems." *Human Factors*, 24, 49-60. doi: 10.1177/001872088202400106

[14] M. R. Endsley (1996). "Automation and situation awareness." In R. Parasuraman & M. Moulou (Eds.), *Automation and human performance: Theory and applications*, 163-181. Mahwah, New Jersey: Lawrence Erlbaum.

[15] M. R. Endsley & E. O. Kiris (1995). "The out-of-the-loop performance problem and level of control in automation." *Human Factors*, 37, 381-394. doi: 10.1518/001872095779064555

[16] J. M. Lockhart, M. H. Strub, J. K. Hawley, & L. A. Tapia (1993). "Automation and supervisory control: a perspective on human performance, training, and performance aiding." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37, 1211-1215. doi: 10.1177/154193129303701802

[17] E. L. Wiener (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft*. NASA Contractor Report No. 177528. Moffett Field, CA: NASA Ames Research Center.

[18] N. B. Sarter, D. D. Woods, & C. E. Billings (1997). "Automation surprises." *Handbook of Human Factors and Ergonomics*, second edition, G. Salvendy (Ed.). New York: John Wiley.

[19] M. R. Endsley (1987). "The application of human factors to the development of expert systems for advanced cockpits." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 31, 1388-1392. doi: 10.1177/154193128703101219

[20] M. R. Endsley & D. B. Kaber (1999). "Level of automation effects on performance, situation awareness and workload in a dynamic control task." *Ergonomics*, 42, 462-492.

[21] V. Riley (1989). "A general model of mixed-initiative human-machine systems." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 33, 124-128. doi: 10.1177/154193128903300227

[22] T. B. Sheridan & W. L. Verplank (1978). *Human and computer control of undersea teleoperators*. Technical Report, Man-Machine Systems Laboratory, MIT, Cambridge, MA.

[23] C. D. Wickens & J. G. Hollands (2000). "Complex systems, process control and automation." In *Engineering Psychology and Human Performance (third edition)*. Prentice Hall, Upper Saddle River, New Jersey, 513-556.

[24] B. Lorenz, F. Di Nocera, S. Rötger, & R. Parasuraman (2001). "The effects of level of automation on the out-of-the-loop unfamiliarity in a complex dynamic fault-management task during simulated spaceflight operations." *Proceedings of the Human factors and Ergonomics Society Annual Meeting*, 45, 44-48. doi: 10.1177/154193120104500209

[25] C. D. Wickens (2008). "Situation awareness: review of Mica Endsley's 1995 articles on situation awareness theory and measurement." *Human Factors,* 50, 397-403. doi: 10.1518/001872008X288420

[26] D. A. Simmon (1998, May- June). "Boeing 757 CFIT accident at Cali, Colombia, becomes focus of lessons learned." Flight Safety Digest, 17(5/6), 1-31.

[27] J. Dino (2008). "Ames Technology Capabilities and Facilities." Retrieved from http://www.nasa.gov/centers/ames/research/technology-onepagers/vms.html

[28] A. L. Singh, T. Tiwari, & I. L. Singh (2009). "Effects of automation reliability and training on automation-induced complacency and perceived mental workload." *Journal of the Indian Academy of Applied Psychology*, 35, 9-22.

[29] E. Mueller, K. D. Bilimoria, & C. Frost (2009). "Effects of control power and inceptor sensitivity on lunar lander handling qualities." AIAA Journal of Spacecraft and Rockets, 48(3), 454-466. doi: 10.2514/1.49276

[30] C. J. Hainley, Jr., K. R. Duda, C. M. Oman, & A. Natapoff (2013). "Pilot performance, workload, and situation awareness during lunar landing mode transitions." Journal of Spacecraft and Rockets, 40, 793-801. doi: 10.2514/1.A32267

[31] J. Wixted & K. Lee (n.d.) "Signal Detection Theory." Retrieved from http://kangleelab.com/signal%20detection%20theory.html

[32] J. B. Lyons & C. K. Stokes (2011). "Human-human reliance in the context of automation." *Human Factors: The Journal of the Human factors and Ergonomics Society*, 54, 112-121. doi. 10.177/0018720811427034

[33] Institute for Homeland Security Solutions. (2010, March). *Trust calibration for automated decision aids*. Durham, North Carolina: McBride, M. & Morgan, S.

BIOGRAPHY

Jessica J. Marquez Ph.D. is a human systems engineer at NASA Ames Research Center. Her expertise is in space human factors and human-computer interaction. Her current projects have focused on software tools for International Space Station (ISS) flight controllers and training simulators. She received her Ph.D. & M.S. from the Massachusetts Institute of Technology from the Department of Aeronautics and Astronautics, and her B.S.E. in Mechanical Engineering from Princeton University.

Margarita Ramirez received a B.S. in Psychology and B.A. in Political Science from the University of Utah in 2008. In 2013, she received her M.S. in Human Factors and Ergonomics from San Jose State University. She has been with the San Jose State University Research Foundation at NASA Ames for two years. Prior to aerospace, she worked at Interface Analysis Associates, a human factors consulting company out of Saratoga, CA.