Evidence Report:

Risk of Inadequate Human-System Integration Architecture

Human Research Program
Human Factors and Behavioral Performance

Approved for Public Release: April 30, 2021

National Aeronautics and Space Administration
Lyndon B. Johnson Space Center
Houston, Texas

CURRENT CONTRIBUTING AUTHORS:

Brian F. Gore, Ph.D.	NASA Ames Research Center
Alonso Vera, Ph.D.	NASA Ames Research Center
Jessica Marquez, Ph.D.	NASA Ames Research Center
Kritina Holden, Ph.D.	Leidos @ NASA Johnson Space Center
Ryan Amick, Ph.D.	KBR @ NASA Johnson Space Center
Donna Dempsey, Ph.D.	NASA Johnson Space Center
Brandin Munson, Ph.D.	University of Houston @ NASA Johnson Space Center

PREVIOUS CONTRIBUTING AUTHORS:

Harsh W. Aggarwal, Ph.D.	Leidos @ NASA Johnson Space Center*
Immanuel Barshi, Ph.D.	NASA Ames Research Center
Maijinn Chen, M.S.	KBR @ NASA Johnson Space Center*
Jolene Feldman, M.S.	San Jose State University Research Foundation @ NASA Ames Research Center
James Garrett, Ph.D.	NASA Johnson Space Center
Maya Greene, Ph.D.	KBR @ NASA Johnson Space Center*
Anikó Sándor, Ph.D.	KBR @ NASA Johnson Space Center*
Sherry Thaxton, Ph.D.	NASA Johnson Space Center
Gordon Vos, Ph.D.	NASA Johnson Space Center

* Left NASA, now at various different institutions and commercial organizations

Acknowledgement

The authors thank Kerry George (KBR/NASA Lyndon Johnson Space Center), for editing this report.

Table of Contents

List of Figures

List of Tables

I. Program Requirement Documents (PRD) RISK TITLE: RISK OF ADVERSE OUTCOMES DUE TO INADEQUATE HUMAN-SYSTEM INTEGRATION ARCHITECTURE (HSIA)

A. Risk statement

As stated in the Human Research Program Roadmap: Given decreasing real-time ground support for execution of complex operations during future explorations missions, there is a possibility of adverse performance outcomes including that crew are unable to adequately respond to unanticipated critical malfunctions or detect safety-critical procedural errors.

B. Context

The National Aeronautics and Space Administration (NASA) Human Research Program (HRP) organizes its research among 5 elements: Human Factors and Behavioral Performance (HFBP), Exploration Medical Capability (ExMC), Human Health Countermeasures (HHC), Research Operations and Integration, and Space Radiation. Elements interface annually with an external panel termed the Human System Risk Board (HSRB) to report progress of risks. Modifications to the content of the risk summaries that derive from changes in the HRP risk research plan require HSRB approval. These include any changes in the high-level deliverables or schedule, and changes or updates in evidence and deliverables that impact baselined LxC risk ratings. In 2016, the Behavioral Health and Performance Element merged with the Spaceflight Habitability and Human Factors Element to create the HFBP Element. The HFBP Element includes several risk areas involving human factors (i.e., risks subsumed under the umbrella of HSIA) and behavioral health (i.e., Sleep, Behavioral Medicine [BMed], Team). In 2018, a new research approach was added, with the cooperation of the HHC and the Space Radiation Elements, to evaluate the potentially synergistic effects of combined, simultaneous exposures to spaceflight hazards that affect the central nervous system (CNS) and operationally relevant behavior and performance; the proposed integrated strategy was dubbed CBS (Central Nervous System/BMed/Sensorimotor) integrated research plan.

In 2019, the Human Factors portfolio was composed of 5 individual risks: vehicle and habitat design (HAB), human-computer interaction (HCI), human-automation and robotics integration (HARI), training (TRAIN), and mission, process and task design (MPTASK). The HSRB requested that these 5 risks be integrated into one new risk termed the HSIA risk with a focus on defining relevant performance outcomes and measures, risk factors, thresholds, and countermeasure guidelines for exploration missions. The HSIA is a conceptual framework addressing the integration of onboard capability and the crew roles and responsibilities necessary to enable the crew to respond effectively and efficiently in the increasingly autonomous mission operations environment. Research in the HSIA risk area is focused on the challenges of exploration missions, i.e., beyond low Earth orbit (LEO). Ground laboratories, analogs of spaceflight stressors, and the International Space Station (ISS) are utilized to answer research questions necessary to ensure that future exploration crewmembers, who will be dealing with many unknowns, communication delays, and increasing autonomy, will be able to adequately perform their tasks and complete their missions. Characterizing the previous Human Factors risks into this one higher level HSIA structure allows the risk to be aligned with measurable outcomes that meet the needs of future operations by identifying gaps in knowledge and countermeasures (CM).

Future deep space missions will present new challenges for crews and increased risks to their performance due to the stress, fatigue, radiation exposure, and isolation that characterizes these missions. In addition, due to distance from the Earth, crews will no longer be able to depend on real-time support from Mission Control Center (MCC) and will have to work increasingly autonomously. Success in this more autonomous environment will depend in part on advanced, onboard automated systems, and on new approaches for training, teaming, and crew selection. Most importantly, it will depend on successfully integrating intelligent vehicle capabilities with crew capabilities.

We must understand the types of tasks that astronauts will likely perform autonomously, and develop human-system integration standards and guidelines for the needed tools and mitigation strategies that ensure success in performing those tasks on an autonomous mission. Enabling a flight crew of 4 to perform the job that has traditionally been done by a ground crew of 40+ will require a fundamental rethinking of crew-vehicle integration and operations. This mode of operation will be very different from the current dependence on frequent real-time direction from a large MCC ground support team of diverse specialties (see Dempsey et al., 2018 for a detailed description of the range of support MCC provides). As outlined by Vera (2019), this new mode of operation will be particularly important when dealing with unanticipated, off-nominal situations. A delay or absence of ground support during unanticipated contingencies can become a significant hazard and can jeopardize the crew and the vehicle if no appropriate onboard capabilities exist to assist with troubleshooting and contingency management. To attain increased autonomy, crewmembers must have complementary and enhanced capabilities that will enable the 4-person flight crew to perform the kind of anomaly response that has previously been done mostly by MCC, and complete these activities in the face of the expected communication delays or unexpected blackouts.

It is important to highlight that complex, autonomous engineered systems will experience unanticipated safety critical malfunctions, due, in part or in whole, to human error or to the impacts of the space environment. Artificial intelligence (AI) and intelligent, autonomous systems can help the human solve problems with complex engineered systems, however, the AI systems will not be able to provide actual problem-solving. Furthermore, reducing the number of crewmembers could increase the risk of non-optimal human system integration (HSI) in unanticipated ways. It is therefore critical that *capabilities be wrapped around the human(s) for safety critical issues to be addressable.*

C. Operational Relevance

The HFBP research Pathway to Risk Reduction has an applied, clear, future-oriented focus on long-duration exploration missions (LDEMs). The goal of the HFBP Element in conducting this human engineering-oriented research, is to transition validated deliverables to operations to make a lasting, positive impact on increasingly autonomous spaceflight crews.

1 EXECUTIVE SUMMARY

The HSIA risk area is focused on the challenges of exploration missions, i.e., beyond LEO. Ground laboratories, ground based analogs of spaceflight stressors (e.g. Human Exploration Research Analog [HERA] as an analog of isolation and confinement during spaceflight), and the ISS are used to answer research questions necessary to ensure that future exploration crews, who will be dealing with many unknowns, communication delays, and increasing autonomy, will be able to adequately perform their tasks and complete their missions. The HSIA is a conceptual framework proposed to address the integration of onboard capability and the crew roles and responsibilities necessary to enable them to respond effectively and efficiently in the increasingly autonomous mission operations environment.

Because the conceptualization of the HSIA is new, no systematic attempt has been undertaken to measure the performance effects associated with reducing the level of ground support that is currently available for mission operations. Consequently, spaceflight evidence for the HSIA risk is lacking, making it difficult to quantify the impact of all the HSIA-related parameters on individual and team-level outcomes. Some degree of causality is discussed in this report (e.g. “communication delay affects ground help,” “distance from Earth affects resupply ability,” etc.). Several HSIA-related CMs, focused in particular areas (e.g. procedure execution support to mitigate procedure errors), are highlighted as potential mitigation options for the overall risk.

The HSIA risk is a young research area for NASA, with limited performance data, but substantial growth opportunities. There is a lack of spaceflight evidence that identifies the requirements for training and crew resources, crew workload capacity, procedure design, expertise pool of the crew, data availability, and vehicle and habitability design that will be necessary to reduce the risk of performance errors in space. Initial evidence for the HSIA risk comes from much of the HAB, HCI, HARI, TRAIN, and MPTASK risks that were part of the NASA HRP Human Factors Portfolio, because they encompass lessons learned from 50 years of spaceflight experience, close to 100 years of aviation, and other ground-based research related to the respective risks. Evidence for the risk of inadequate HSIA comes primarily from examining data from investigations, focus groups, case studies, crew reports, and accident investigations, however, some evidence is available from recent spaceflight studies and from MCC reports. Because the number of commercial, military, and private flights each year far exceeds the number of spaceflights, much of the evidence comes from aviation research and from accident reports from other domains. In addition, some of the evidence consists of summaries of subjective experiences and non-experimental observations or comparative, correlational, and case studies. Some evidence in this report is derived from the Flight Crew Integration (FCI) ISS crew comments database and the Shuttle external crew reports. Although summaries of ISS and Shuttle crew comments are presented as evidence, the FCI ISS crew comments database is protected and not publicly available due to the sensitive nature of the attributable data it contains. Data are also presented from the Crew Office approved Space Shuttle Crew Reports. These reports are not publicly available.

2 INTRODUCTION

Future deep space missions will present new challenges for crews, and increased risks to human performance due to the stress, fatigue, radiation exposure, and isolation that characterizes these missions. In addition, crews will no longer be able to depend on real-time support from MCC due to distance from the Earth, and they will have to work increasingly autonomously, performing critical tasks that were previously carried out by ground experts. Success in this more autonomous environment will depend on advanced, onboard automated systems, and on new approaches for training, teaming and crew selection. Most importantly, it will depend on successfully integrating intelligent vehicle capabilities with crew capabilities.

Ground laboratories, analogs, and the ISS are utilized to answer research questions that will ensure future exploration crews, who will be dealing with many unknowns, communication delays, and increasing autonomy, can adequately perform their tasks and complete their missions. The HSIA is a conceptual framework proposed to address the integration of onboard capability and the crew roles and responsibilities necessary to enable crews to respond effectively and efficiently in the required increasing autonomous mission operations environment.

Enabling an exploration crew of 4 to perform the job that has traditionally been done by a ground crew of 40+ will require a fundamental rethinking of crew-vehicle integration and operations. This mode of operation will be very different from the current dependence on frequent real-time communications with a large MCC ground support team of diverse specialties (see Dempsey et al., 2018 for a detailed description of the range of support MCC provides). As outlined by Wu and Vera (2019), this new mode of operations will be particularly important when dealing with unanticipated, off-nominal situations. To attain increased autonomy, crewmembers must have complementary and enhanced capabilities that will enable the 4-person flight crew to perform the kind of anomaly response that has previously been done mostly by MCC, and complete these activities in the face of the expected communication delays or unexpected blackouts.

It is important to highlight that all complex, engineered systems, autonomous or not, will experience unanticipated safety critical malfunctions due, in part or in whole, to human error and/or to the impacts of the space environment. Although AI and intelligent/autonomous systems can help the human solve problems that arise with complex engineered systems, the AI systems will not be able to provide actual problem-solving. Furthermore, reducing the number of crew in any system has the potential to increase the risk of non-optimal human HSIA in unanticipated ways. It is therefore critical that c*apabilities be wrapped around the human(s) for safety critical issues to be addressable.* To further mitigate the risk, specific factors have been identified (e.g., autonomy, anomaly detection, and response, etc.) as highly relevant to LDEMs, and research will focus on understanding the impact of these factors on LDEMs.

3 SOURCES OF EVIDENCE

This evidence report documents the current state of knowledge about the risk and its mitigation and interactions with other spaceflight risks. It includes relevant and applicable evidence from sources including HRP-funded research, previously published data, international partner-funded research, research from related fields, and unpublished ideas (e.g., paradigms, perspectives). Evidence includes space-related evidence (where available) and ground-based evidence.

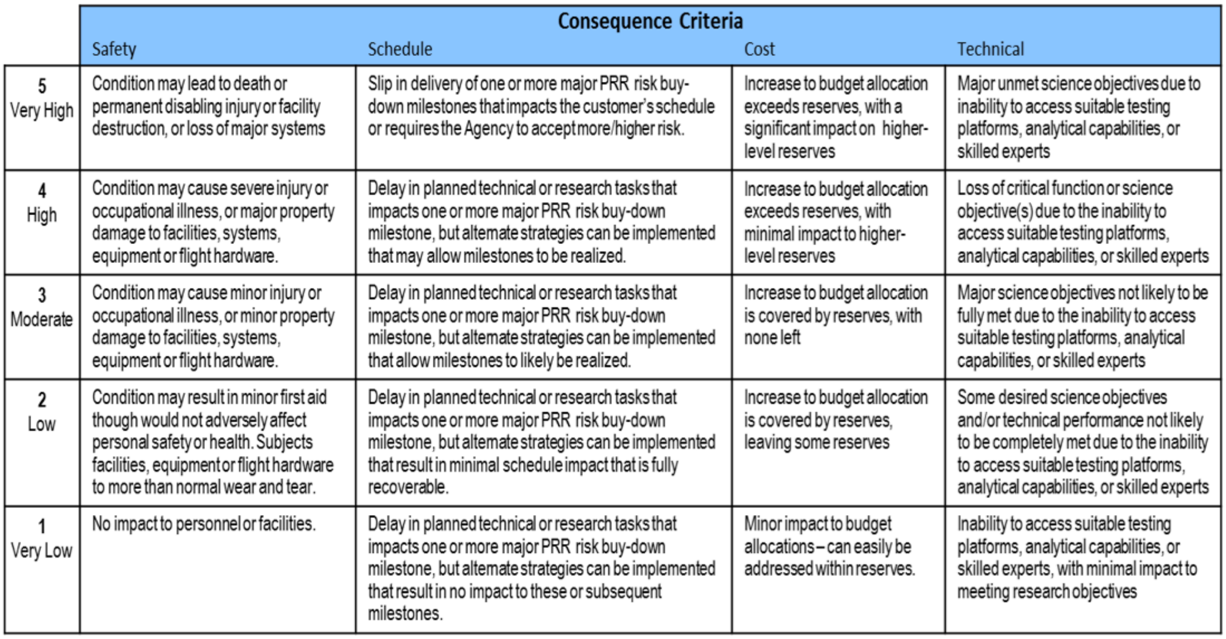
There are 5 primary hazards for human spaceflight; (1) radiation; (2) isolation and confinement; (3) distance from Earth; (4) altered gravity; and (5) hostile/closed environment. HSIA is a complex, integrated risk that is impacted by 3 of the 5 primary hazards (2, 3, and 5) and includes risks of performance impacts due to incompatible vehicle and habitat designs, inadequate integration of humans with automation and robotics, inadequate integration of humans and computer systems, inadequate training, and inadequate design of processes and tasks for responding to anomalies in an autonomous environment.

4 LEVELS OF EVIDENCE

The purpose of risk analysis is to characterize the significance of risks once they have been identified and validated. An evaluation of risk probability and impact/severity is used to establish the significance relative to other risks. The likelihood and consequence (LxC) score for an HRP programmatic risk is qualitatively assessed based on the assumption that available CM and mitigations are taken. A scale of 1 to 5 is used, where 1 represents the lowest likelihood or least consequence, and 5 represents the highest likelihood or consequence.

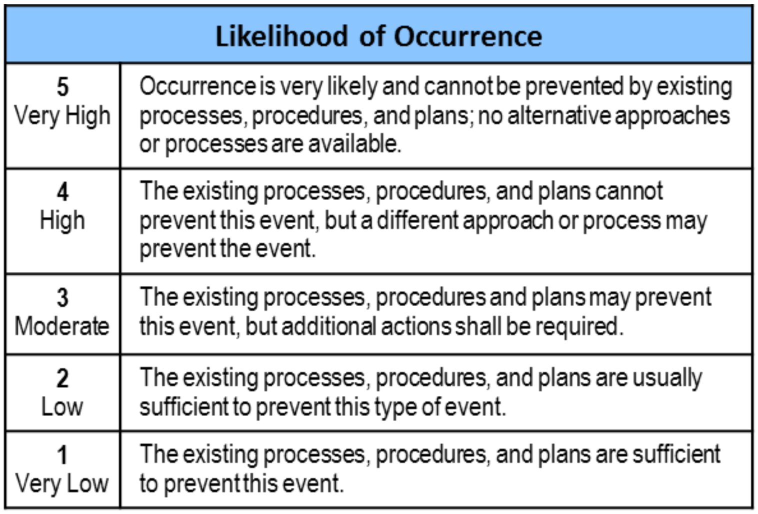
For each risk, the risk owner clearly defines the consequence, determines the criterion (safety, schedule, cost and technical) to which it belongs and assigns a score based on the definitions for that category (Table 1). If a risk is applicable to multiple criteria, the Risk Owner chooses the highest consequence score assigned. Next, the risk owner evaluates the likelihood of the selected consequence and applies a score based on the likelihood scale definitions (Table 2). The rationale and the drivers behind the selected scores are documented, and the uncertainty noted as appropriate.

Table 1. Consequence Criteria and Scale



	Consequence Criteria			
	Safety	Schedule	Cost	Technical
5 Very High	Condition may lead to death or permanent disabling injury or facility destruction, or loss of major systems	Slip in delivery of one or more major PRR risk buy-down milestones that impacts the customer's schedule or requires the Agency to accept more/higher risk.	Increase to budget allocation exceeds reserves, with a significant impact on higher-level reserves	Major unmet science objectives due to inability to access suitable testing platforms, analytical capabilities, or skilled experts
4 High	Condition may cause severe injury or occupational illness, or major property damage to facilities, systems, equipment or flight hardware.	Delay in planned technical or research tasks that impacts one or more major PRR risk buy-down milestone, but alternate strategies can be implemented that may allow milestones to be realized.	Increase to budget allocation exceeds reserves, with minimal impact to higher-level reserves	Loss of critical function or science objective(s) due to the inability to access suitable testing platforms, analytical capabilities, or skilled experts
3 Moderate	Condition may cause minor injury or occupational illness, or minor property damage to facilities, systems, equipment or flight hardware.	Delay in planned technical or research tasks that impacts one or more major PRR risk buy-down milestone, but alternate strategies can be implemented that allow milestones to likely be realized.	Increase to budget allocation is covered by reserves, with none left	Major science objectives not likely to be fully met due to the inability to access suitable testing platforms, analytical capabilities, or skilled experts
2 Low	Condition may result in minor first aid though would not adversely affect personal safety or health. Subjects facilities, equipment or flight hardware to more than normal wear and tear.	Delay in planned technical or research tasks that impacts one or more major PRR risk buy-down milestone, but alternate strategies can be implemented that result in minimal schedule impact that is fully recoverable.	Increase to budget allocation is covered by reserves, leaving some reserves	Some desired science objectives and/or technical performance not likely to be completely met due to the inability to access suitable testing platforms, analytical capabilities, or skilled experts
1 Very Low	No impact to personnel or facilities.	Delay in planned technical or research tasks that impacts one or more major PRR risk buy-down milestone, but alternate strategies can be implemented that result in no impact to these or subsequent milestones.	Minor impact to budget allocations – can easily be addressed within reserves.	Inability to access suitable testing platforms, analytical capabilities, or skilled experts, with minimal impact to meeting research objectives

Table 2. Likelihood Scale



Likelihood of Occurrence	
5 Very High	Occurrence is very likely and cannot be prevented by existing processes, procedures, and plans; no alternative approaches or processes are available.
4 High	The existing processes, procedures, and plans cannot prevent this event, but a different approach or process may prevent the event.
3 Moderate	The existing processes, procedures and plans may prevent this event, but additional actions shall be required.
2 Low	The existing processes, procedures, and plans are usually sufficient to prevent this type of event.
1 Very Low	The existing processes, procedures, and plans are sufficient to prevent this event.

The resulting LxC score is then plotted on a 5 x 5 risk matrix shown in Figure 1. The LxC score is associated with the color of the cell within which it falls.

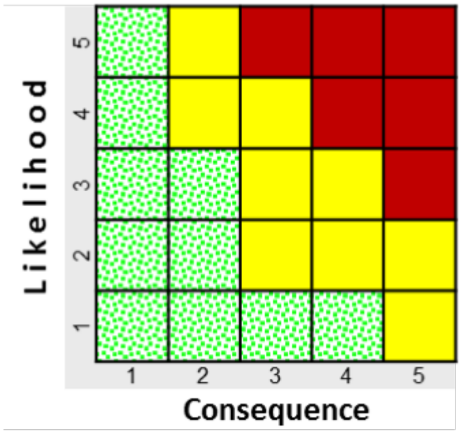


Figure 1. Risk LxC Matrix

Initial risk information is entered into the Johnson Space Center (JSC) Integrated Risk Management Application (IRMA) tool, which maintains the official risk records of JSC programmatic risks.

In certain instances, as determined by the Program Director, it may be necessary to analyze a set of risks as an aggregate risk. This type of risk is a view of the set of risks as they contribute to a broader consequence outcome. The same LxC scales apply to its assessment.

Risk Planning

As a new risk is identified and analyzed by the Risk Owner, the next step is to determine whether the responsibility for its mitigation is to be retained by the Risk Owner, delegated to another Risk Owner or transferred to another program, project, or organization that is better suited to address the risk.

Once the appropriate responsibility has been determined, the Risk Owner considers available resources and considers the best approach to take:

- Mitigate: Specific activities are performed to reduce the risk in order to minimize the probability of occurrence and/or severity of the consequence. Within the approved mitigation strategy, each mitigation step or set of steps will have an LxC score that traces the path to an LxC goal that makes the risk acceptable to the program. The criteria by which to evaluate the success of each mitigation step are identified. A brief description of the resources required for implementing each mitigation step is also documented.
- Investigate (Research): Additional information is gathered within an agreed upon period of time to determine how the risk can be evaluated for its LxC or whether there is a better plan forward.

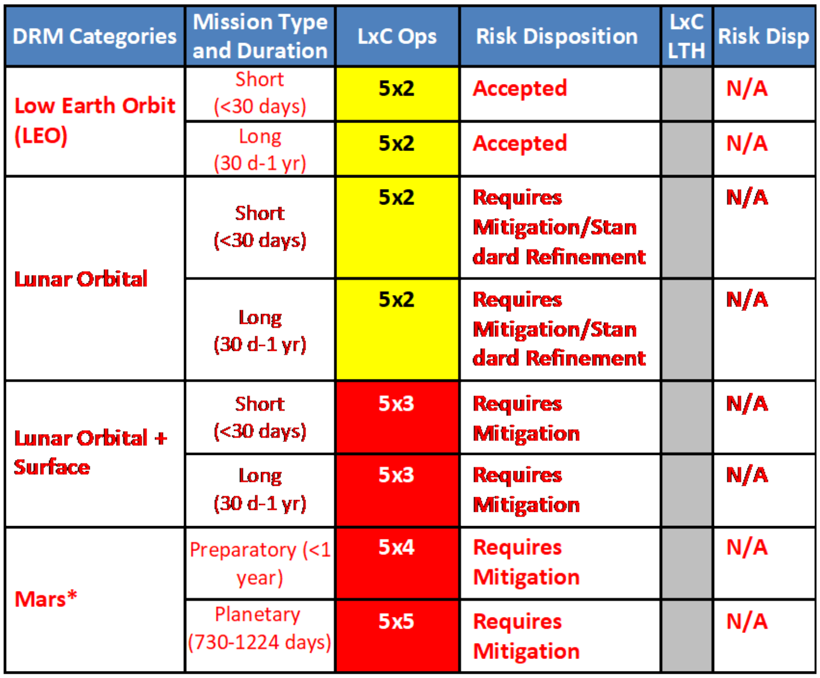
- Watch: The risk and its attributes are monitored for early warning of critical changes in impact, probability or other aspects with agreed upon triggers or thresholds that would necessitate action.
- Accept: HRP is willing to accept the consequences and the associated likelihood, and resources are not allocated for managing the risk. Accepted risks will be reviewed periodically to ensure that the acceptance rationale is still valid, and the review cycle will be noted in the acceptance rationale.

The mitigation strategy for a risk should consider relationships with other risks that are defined in terms of scope or potential shared activities or issues. These relationships are clearly noted in the risk record and factored in the evaluation of the progress of risk reduction. This evaluation may also identify cross-cutting risks, which are risks generally applicable to multiple HRP efforts, with attributes and impacts found in multiple levels of the organization or in multiple organizations within the same level.

When necessary, contingency plans are made to reduce the severity of impact should the adverse event, as identified by the risk, occur.

The HSIA risk has the following risk profile (Table 3). The risk disposition for the HSIA per Design Reference Mission (DRM): Operations for LEO missions accepted based on current CMs and availability for rapid evacuation to Earth, timely resupply, and regular real-time communications. For short-duration lunar orbital and lunar orbital + surface missions, risk disposition is classified as *Standard Refinement* and possibly requires mitigation due to new, incremental technology adaptation. For missions of greater than 30 days in lunar orbit, lunar surface missions, and Mars missions, new standards, guidelines and other mitigation are required due to greater unknowns, long mission durations, limited evacuation/resupply options, and decreasing real-time support. Mission systems must support increasingly independent crew operations possibly requiring increased automation, robotics and intelligent technologies. Thus, dispositions for these DRM are classified as *Requires Mitigation*. Long Term Health (LTH) considerations for Mars missions likely require mitigation to minimize impact on the crew. Mitigation strategies require further research.

Table 3. HSIA LxC Risk Disposition (updated 01/2021)



DRM Categories	Mission Type and Duration	LxC Ops	Risk Disposition	LxC LTH	Risk Disp
Low Earth Orbit (LEO)	Short (<30 days)	5x2	Accepted		N/A
	Long (30 d-1 yr)	5x2	Accepted		N/A
Lunar Orbital	Short (<30 days)	5x2	Requires Mitigation/Standard Refinement		N/A
	Long (30 d-1 yr)	5x2	Requires Mitigation/Standard Refinement		N/A
Lunar Orbital + Surface	Short (<30 days)	5x3	Requires Mitigation		N/A
	Long (30 d-1 yr)	5x3	Requires Mitigation		N/A
Mars*	Preparatory (<1 year)	5x4	Requires Mitigation		N/A
	Planetary (730-1224 days)	5x5	Requires Mitigation		N/A

Primary CMs for the risk of inadequate HSIA include prevention, monitoring, and intervention. Prevention CMs require extensive simulation-based assessment of operational tools and procedures, technology solutions and human-centered automated systems, training and teaming, using time-delayed flight controller monitoring and support. System requirements and designs need to emphasize contingency management and maintainability/reparability of vehicle as well as for supporting crew performance. The Monitoring CMs *require* onboard sensor networks for vehicle and crew state. Intervention CMs require designing systems for updatable software, just-in-time training, HSIA that enables onboard problem-solving and complex procedure execution, including new training and teaming methods, and human centered design for automated systems.

The primary focus of the HSIA is to prepare and equip crewmembers with skills and tools for operating autonomously in response to unexpected and off nominal events, thus leading to improved safety, efficiency, and mission success. Human-system designs need to be centered on the human's capabilities and limitations during LDEMs. Safety and mission critical anomalies will be inevitable on LDEMs, and delays and interruptions in communication with experts on Earth- drives the requirement that crew resolve these anomalies on their own. Reliable, onboard capabilities will be needed to support the crew—not only capabilities to resolve anomalies, but also to promote situation awareness (SA) and to reduce workload. Unfortunately, the present level of technological advancement limits the range of achievable intelligent support (Wu & Vera, 2019). The Human Capabilities Assessment for Autonomous Missions (HCAAM) projects was devised to identify how technology may benefit crew SA and enhance crew trust in the intelligent systems.

Spaceflight missions require crewmembers to perform a wide variety of tasks under dramatically different conditions: 1-g, hypergravity, microgravity, unsuited, suited, and pressurized under nominal and off nominal operational environments. A mapping of the 5 primary risks of human spaceflight to the HSIA contributing factors and the likely CMs for these risks areas has been presented to the HSRB in a directed acyclic graph (DAG). The HSIA DAG representation starts with hazards (yellow boxes; e.g. distance from Earth/hostile closed environment) associated with deep space crewed exploration, the contributing factor (blue boxes; e.g. autonomous operations/anomalous events) progressing to the consequences if the hazard is not mitigated (red boxes; loss of crew/loss of vehicle) as a function of the distance from Earth. The HSIA DAG represents causality using arrows and dashed lines ("commination delay affects ground help," "distance from earth affects resupply ability," etc.). The anticipated CMs that will be needed to mitigate the overall risk are outlined in the green boxes in Figure 2, focused in particular areas (e.g., communication tools, procedure design/execution support to mitigate procedure errors, crew training, among others).

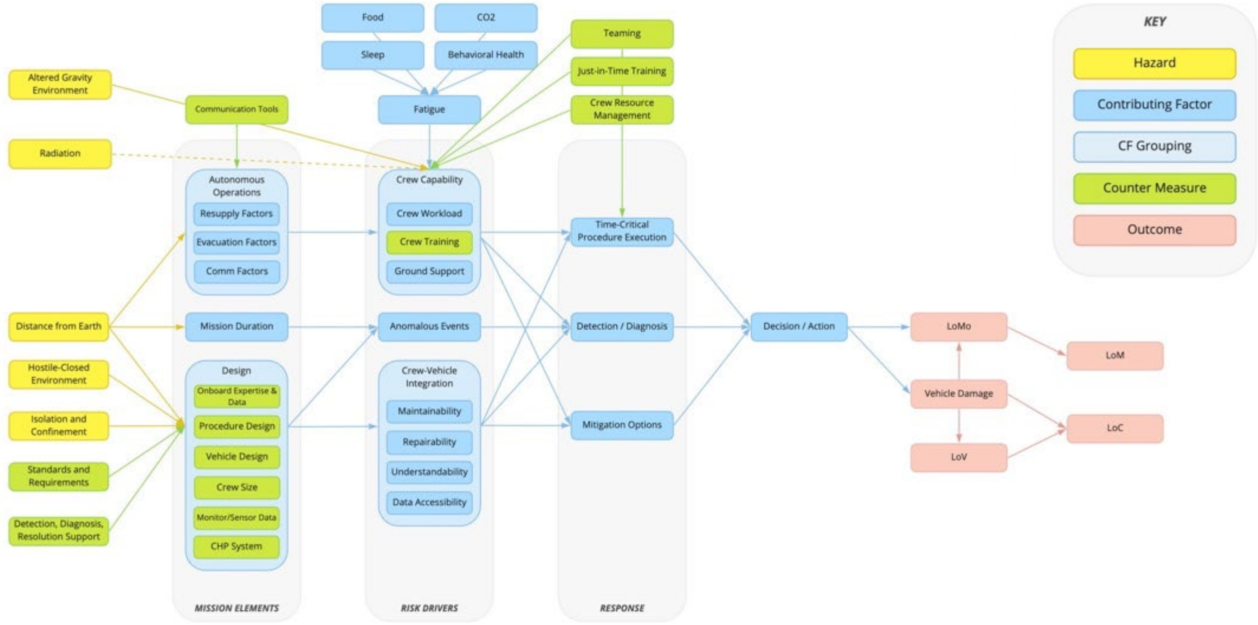


Figure 2. HSIA Directed Acyclic Graph with Expected Countermeasures (as of January 2021)

The HSIA DAG is still maturing but focuses on the crew capabilities that must be available on board for mission success. These include the capability to understand and perform tasks; detect and recover from errors; maintain and repair hardware, software, and procedures; and troubleshoot and diagnose problems, particularly problems that were unanticipated and for which no training was provided. Critical malfunctions requiring crew/MCC management have occurred at an average rate of 1.7 times per year over the lifetime of the ISS, and 3-4 times per year during the "burn in phase" for the vehicle. These averages do not include extra vehicular activity (EVA) data, which greatly increase the incident rate. Prior experience from the Apollo program showed 10 of 11 crewed missions experienced significant anomalies that required the crew to rely heavily on MCC expertise in real-time. These failure patterns are in line with those observed in other complex engineered systems (e.g., oil rigs, launch systems, commercial aviation, etc.).

Other critical HSIA related data driving the need to reduce the risk areas include the following:

- Likelihood of high consequence problem of uncertain origin > 10 % in 30-day mission.
- Ground support may be less timely and less comprehensive for Moon and Mars DRMs. Crew will have to perform more problem solving and complex operations without 24/7 support and using existing onboard supplies.
- Response to anomalies (e.g., medical issues) are anticipated to be most difficult, most important tasks undertaken by Earth independent crews.
- Anomaly response requires human intervention; independent capability needs to exist locally, tailored to size and capacity of team.
- Increased complexity of mission systems and interaction leads to increased likelihood of adverse events.
- HSIA-related anomalies during Moon missions are a significant proportion of total anomalies reported
- HSIA issues in Expedition 1 were many and common and similar issues are still arising 20 years later.
- Level of ground support experience/expertise cannot be replicated wholesale on board.
- Problem solving skills and training are specific to the role but not domain
- Lines of inquiry in anomaly resolution can be supported but not invented by machines.
- HSIA must provide problem-solving support.
- Decision domains vary as anomalies unfold; each domain requires different onboard decision support for crew.
- HSIA has evolved throughout the spaceflight era but has not fundamentally changed.

- Next generation HSIA will require a revolutionary shift.

To support HSIA, it is critical to consider the kinds of systems and responsibilities being supported by the crew, the HSIA concept of operations and operational contexts, stressors that impact the crew's operational performance, and measures and metrics that can be used to quantify human performance in relation to the HSIA modifications.

HSIA Needs Assessment

The first step toward refining the HSIA is identifying the functions the vehicle habitat needs to support and the knowledge and expertise the crew need to possess. These data, tools, knowledge and expertise used in the resolution of historical safety-critical off-nominal incidents can then be validated against observations of simulation training conducted in the -Space Station Training Facility and during flight controller interviews. High Priority events—events that could result in loss of crew (LOC) or loss of mission (LOM) occurred as frequently as once every 2 months during the first 2 years of ISS operations (Wu & Vera, 2019). The items for investigation (IFI) database includes 67 high priority IFIs from 2002 to 2019, 35 of which are associated with malfunctions in vehicle subsystems, including active thermal control, electrical power, ECLSS, guidance, navigation and control, and structures subsystems (Panontin, et al., 2021). Of these 35 anomalies, 18 required urgent responses. These types of urgent responses occurred about once per year, although the distribution is heavily skewed toward early operational years (e.g., 3 occurred in 2002, 3 in 2004).

Although valuable, the data contained in the IFI reporting system are generally high-level summaries of the investigations and resolution steps. Further analysis of the events that required immediate response by ground personnel will determine the critical tasks that the crew must perform autonomously during future LDEMs. As many as one-third of high-level IFIs may have required immediate response from ground control. To characterize the HSIA correctly, research teams observe the deliberations of anomaly resolution teams to gain a deeper understanding of how urgent diagnosis and resolution are currently done and what products are used to aid the process. Ultimately this effort will identify the data and expertise an autonomous crew will require to respond to an anomaly during a LDEM, and can be used to support the development of performance operating limits (POLs). The context of operations surrounding the decisions taken and the actions made is integral to the HSIA.

HSIA Concept of Operations (ConOps)

Holden et al. (2019a) identified a suite of time-critical and complex exploration mission tasks that may advance the development of human-integration ConOps for an autonomous crew. The HSIA ConOps presents a progressive maturation of anomaly response and execution of complex operations technology (e.g., data systems, training, human-automation teaming strategies, collaboration support, task support, planning support, work management support, knowledge-based decision support, etc.). Feedback from relevant subject matter experts (SMEs) is being obtained through a ConOps Working Group in early 2021. This work enables systems engineers-to develop detailed system requirements and the HRP to establish a high-level basis for future research needs.

HSIA, Anomaly Response, Environmental Operations and Stressors (CO_2)

Due to power constraints, CO_2 levels on the ISS are higher than terrestrial levels: CO_2 levels are around 2500 parts per million on ISS (Simon et al., 2011), and it is anticipated that exploration vehicles and habitats will have similar levels. Levels of CO_2 above 2.5 mmHg are associated with increased incidence of headaches (Law et al., 2014) and changes in the mood and performance of crewmembers, although performance decrements from elevated CO_2 levels are not well established. As a result, permissible exposure limits (PELs) are critical to recognize and link to HSIA. Beard (2020) developed an anomaly response framework that depicts how spaceflight and task-related stressors, training, team behavior, goals and metacognitive processes can influence crew anomaly response and how an intelligent, learning system could help mitigate the HSIA risk. To identify how elevated CO_2 may impact a self-reliant crew's anomaly response, Beard (Beard, 2020) traced the relationship between CO_2 levels and the cognitive and motor abilities required to perform critical mission tasks (Holden et al., 2019a) and to detect, identify, and respond to anomalies, and concluded that elevated CO_2 could exacerbate the HSIA risk. Reliable, onboard capabilities will need to support the crew—not only to detect and resolve anomalies, but also to promote situation SA and to reduce workload.

HSIA Contextual Inquiry

A semi-structured interview in the field of human-systems integration, design, and architecture (termed a contextual inquiry methodology) was completed. This process lay the foundation for the key drivers of the problem-solving process (e.g., anomaly analysis and resolution, contingency planning, etc.). Observations of current mission control processes, and relevant mission simulations of responses to anomalous events was also completed to characterize the roles, tools, data flows, teaming, and training elements to understand how tasks are actually performed in the current settings. This task also consisted of site visits to NASA JSC to observe ISS mission operations and mission operations simulation (Vera et al., 2019). Interviews conducted with Flight Controllers, Flight Directors, and former astronauts indicate that, for current ISS operations, when an unanticipated anomaly with LOC or LOM potential occurs where the nature of the problem is not understood (the sorts of anomalies that can be found in the IFI data system), aa large team, including Front Room, Back Room, and Mission Evaluation Room Controllers (up to about 40 people) respond immediately. In these situations, a "Team Four" is often called in to support the anomaly resolution process. It takes around 3 to 4 hours to invoke a Team Four.

HSIA and Lunar Lander

Decreasing real-time ground support during LDEMs will require a higher level of crew autonomy and consequently crew's readiness to perform. An assessment was completed of potential impacts of the Artemis Program's proposed Human Landing System (HLS) on crew's readiness to perform in relation to HSIA. Apollo crews' reported experience and empirical research findings are used to consider potential impacts from qualitative and quantitative perspectives. On the whole, Apollo astronauts sustained minimum impacts on their sensorimotor and cognitive functions from the Apollo lunar module's acceleration profiles and successfully completed Moon landing and lunar surface operations. However, the major operational differences expected for the Artemis missions (use of Gateway orbit and South Pole of the Moon landing) combined with insufficient knowledge of thresholds for human sensorimotor function and cognitive performance and anticipated crew roles and tasks, make it impossible to accurately assess the actual impacts on operational performance, research necessary for establishing POLs.

5 HUMAN FACTORS MEASURES AND METRICS TO ASSESS HSIA

The evidence for risk of inadequate HSIA depends largely on measures of performance. Wenzel (2019) provides recommendations on measures to assess the risks to human performance that are related to HSIA. Because selection and interpretation of appropriate human performance measures depends on contextual factors such as the specific task, operation, or human-machine system involved, Wenzel recommended:

- Metrics that measure human performance should be developed as an inherent part of the technological infrastructure of spaceflight analogs such as HERA and onboard future crewed spacecraft. For example, software (e.g. ProX) could be developed that supports execution of a procedure, measures performance, and validates the usefulness of particular adaptive, in-mission training and performance support tools.
- Baseline performance levels and variability in normal performance of critical tasks should be established, as should criteria for determining when performance is significantly worse or better; these criteria will be skill or task-dependent and may depend on individual differences among crewmembers.
- To the extent possible, metrics should be obtained unobtrusively. If measured as an inherent part of onboard systems, accuracy and time measures may be made unobtrusive. However, such metrics will need to be tailored for individual tasks and skills.
- Physiological metrics or oculometrics could be unobtrusive metrics, although further work is needed to definitively demonstrate consistent physiological signatures for various human behavioral experiences such as workload or SA. The consistency of

such metrics will likely be complicated by a variety of factors such as fatigue, stress, and individual differences in experience or proficiency.

- The judicious use of more overt metrics in the form of surveys and crew reports may remain a necessity because it is unlikely that unobtrusive measures will monitor some aspects of the crew's performance, health, and safety. These may include survey-based measures of workload (NASA task load index [TLX], Bedford workload scale), SA (situation awareness global assessment technique [SAGAT]; situation present assessment method [SPAM]; SA rating technique [SART]), trust in automation (trust in automated system scale [TASS]; human-computer trust scale [HCTS]), and habitability (iPad-based Space Habitability Observation Reporting Tool, iSHORT; iPad-based Question & Answer, iQ&A). However, some of these metrics may be more appropriate during development of spacecraft systems rather than as an onboard routine.
- Tools that assess sensorimotor abilities may also useful during LDEMs. In addition to providing baseline measures of unimpaired sensorimotor abilities, these tools can provide insight into crew cognitive states and/or limitations induced by factors such as fatigue, sleep deprivation, hypoxia, hypercapnia, and drug-induced effects. For example, the psychomotor vigilance test (PVT), the cognition test battery, and the comprehensive oculomotor behavioral response assessment (COBRA), are reliable, relatively brief, and may be easy to administer during LDEMs.
- Operational tasks could be used to assess the optimum relationship between crew training, performance, and actual operations, and could identify performance measurements that assess and improve astronaut interaction with systems. Given sufficient resources, it may be possible to expand the research capabilities of projects currently in use or in development such as the robotics on-board trainer (ROBoT-r), the emergency onboard training simulator, and the Pro-X adaptive user interface infrastructure, for procedure execution.

Table 4 presents recommended measures for operational evaluations and how these approaches have been used in research. Some of these methods and measures will be highlighted in the following evidence section.

Table 4: Recommended measures for operations vs. research & development. Please see Wenzel (2019) for details on individual measures.

Recommended Measure	Operational Studies, e.g., on ISS	Research & Development
Accuracy*	% correct, other accuracy measures as appropriate to task	% correct, other accuracy measures as appropriate to task
Mean Absolute Error (MAE), Root Mean Square Error (RMSE)*	MAE, RMSE for tracking or similar tasks	MAE, RMSE for tracking or similar tasks
Time*	Response or reaction time, time on task, task/step completion time, other time measures as appropriate to task	Response or reaction time, time on task, task/step completion time, other time measures as appropriate to task
Workload	Accuracy and time used as converging operations for primary, secondary tasks*; use with subjective measures like Bedford Workload Scale that requires less time & effort to administer than NASA-TLX	Accuracy and time used as converging operations for primary, secondary tasks*; use with subjective measures like NASA-TLX, considered more reliable and well-validated
Situation Awareness (SA)	SART, subjective measure that requires less time & effort to administer	SAGAT, SPAM: objective measures, both are probe techniques for high fidelity sims that require more time & effort to administer
Trust in Automation	Objective measures of automation compliance, time, accuracy* combined with single question rating of trust	Objective measures of automation compliance, time, accuracy* combined with Trust in Automated System Scale (TASS) or Human Computer Trust Scale (HCTS)
Assessment Tools for Sensorimotor Ability	PVT+, Cognition Test Battery, COBRA (likely pre- and post-task use)	PVT+, Cognition Test Battery, COBRA (likely pre- and post-task use)
Operational Tasks	Tasks such as ROBoT-r and Pro-X adaptive procedure platform still in development	Tasks such as ROBoT-r and Pro-X adaptive procedure platform still in development
Habitability	iSHORT, iQ&A	iSHORT, iQ&A

* Assumes measures are collected as unobtrusively as possible in that they have been integrated into the technological infrastructure.

The notion of error rate is particularly important, as exemplified in the National Transportation Safety Administration Board (NTSB) report from the SpaceShip Two (SS2) in-flight breakup during its test flight (NTSB, 2014a). The NTSB outlines that lack of human factors guidance for commercial space operators contributed to the un-commanded feather extension on the SS2 vehicle. The development process did not include opportunities to identify design and operational factors that could have mitigated the catastrophic consequences of a single human error during a high workload phase of flight. To prevent a similar situation from recurring, commercial space operators should fully consider human factors when designing and operating space vehicles. However, because commercial spaceflight is an emerging industry, no guidance currently exists that advises commercial space operators to obtain human factors expertise, consider human error in hazard analyses, ensure that hazard analyses avoid or adequately mitigate single-point failures, and ensure that flight crews are aware of known catastrophic hazards that could result from a single human error. This backdrop sets the stage for the integrated nature of the HSIA.

Society of Automotive Engineers (SAE; 2019) HSI as "the management and technical discipline of planning, enabling, coordinating, and optimizing all human-related considerations during system design, development, test, production, use and disposal of systems, subsystems, equipment and facilities." These systems often involve time-critical tasks, that is, tasks that typically have a specific onset time and a specific time by which the task needs to be completed. Together these define a window of opportunity for the action to take place. For such systems, the dynamic interactions among system elements are often critical for the human to control the system. HSIA refers to the architecture, or framework that addresses the integration of onboard capability and crew roles and responsibilities necessary to enable crew to respond effectively and efficiently in an increasingly autonomous mission operations framework. Because the HSIA will require that crew operate in an autonomous fashion in response to unforeseen events, much of the crew performance will require humans to interact with complex computer systems, and with other forms of automation, controls, displays, and robotic systems to address critical activities that will sustain crew safety and life. Once the crew is far enough from Earth, the manner that communication occurs changes fundamentally in terms of cadence, delay, and bandwidth. When the communication ability changes, so does the manner that the crew undertake resolution activities. The crew will rely increasingly on onboard problem-solving abilities as well as tools to support them. These CMs reduce the risk of a LOC or LOM.

Risk of Inadequate HSIA and Vehicle Habitat Design Factors

To promote safe and efficient human performance during space missions, it is important to consider in the design process the effects of microgravity, acceleration, vibration, and other environmental conditions. It is also important to consider human capabilities and limitations with respect to the interface design of the space habitats, vehicles and other systems, and how these interactions and interfaces may change on LDEMs. When these things are not considered, there is a risk that the vehicle habitat design will be incompatible with human capabilities and limitations. This applies to all space habitats and transit vehicles. Examples of short-term effects due to this risk include overexertion, difficulty in reading a checklist due to spacecraft vibrations or inadequate lighting, high temperature in a module due to inefficient co-location of habitability-related hardware and excessive activities, difficulty donning a suit due to inadequate habitable volume, and difficulties in communicating with fellow crewmembers due to high levels of noise in the cabin. Performance-related inefficiencies may include unnecessary translations between workstations to complete tasks, increased task completion time due to difficulty in accessing equipment, and lack of restraints for performing tasks requiring stability. Examples of long-term effects include ergonomic-related cumulative trauma disorders from repetitive motions and/or sustaining awkward postures, insufficient workspace clearances resulting in frequent over-exertions, and suit hardware requiring sustained performance at maximal levels, and long-term effects could also include permanent hearing loss. Interacting with a vehicle habitat that does not accommodate the crew along all anthropometric ranges and does not consider human physical and cognitive capabilities, limitations, and how these may change during long-duration spaceflight, could lead to injuries, crew frustration, and/or mission failure.

The HSIA-related design factors for the habitat include (1) anthropometric and biomechanical limitations, (2) visual environments, (3) vibration and g-forces, (4) noise interference, (5) seating, restraints and personal equipment, (6) visibility and window design and placement, and (7) the habitat volume and layout of the vehicle and the minimum net habitable volume (NHV) to accomplish these tasks to prevent and reduce injuries and inefficiencies. During the development, test, and evaluation phases, it will be critical to perform activities such as population analysis, digital modeling, simulations, and that astronauts with previous spaceflight experience participate in human-in-the-loop (HITL) evaluations to assess and prevent incompatibilities between the human and the vehicle habitat. The use of mockups and simulators with an appropriate level of fidelity to accurately represent the configuration of the vehicle habitat layout will be key to designing for microgravity and reduced-gravity environments. This evidence report introduces habitability and each contributing factor; describes spaceflight evidence for each contributing factor; describes ground-based evidence for each contributing factor; discusses relevant modeling and simulation considerations; and considers risk in the context of operational scenarios.

To design a work environment that can accommodate crew capabilities and limitations, these capabilities and limitations must be well understood and documented for all the mission phases. The vehicle habitat designers must better understand the strengths and weaknesses associated with performing tasks in microgravity and reduced gravity environments, as well as the minimum NHV required to accomplish these tasks to prevent and reduce injuries and inefficiencies. Mockups and simulators with an appropriate level of fidelity to accurately represent the layout configuration of the vehicle habitat will be key to properly replicating the vehicle habitat experienced in microgravity and reduced gravity environments. These mockups can provide insight into design, and later continue to be used for design validation and training purposes. Another challenge is verifying requirements during development when many systems are being developed concurrently—we must rely, in part, on high-fidelity computer models. For example, spacesuits in a range of anthropometric sizes and a realistic acoustic environment are generally not available for testing during development. The lack of adequate verification of the requirements during development may result in noncompliance with the requirements and adverse effects on the crew during operations.

Finally, during the operational phase, lessons learned regarding the use of the vehicle habitat need to be captured so that ergonomic and environmental issues and impacts are identified and validated, and CMs and interventions can be applied. Implementing this into the HSIA umbrella will require development and use of habitability and ergonomic assessment metrics, tools and methodologies.

Anthropometric and Biomechanical Limitations

When the size, strength, dexterity, or mobility of the human body are restricted this can create an unsafe situation. Although flight accidents solely due to anthropometry and biomechanical limitations are somewhat difficult to quantify, studies have shown that poor consideration of physical body, shape, size, and exertion capabilities, as well as inadequate clearances around the operator and the surrounding work interfaces could lead to severe injuries (e.g. repetitive stress injuries or amputation) as well as loss of life (Konz & Johnson, 1999).

Visual environments

When factors such as weather, haze, darkness, dust, and smoke inside or outside the vehicle restricts the vision to a point where normal duties are affected, this can create an unsafe condition. Poor visibility conditions likely contribute to error, injury, or poor task performance. Appropriate lighting is critical in a spacecraft because visual perception is the crewmember's primary method of obtaining information about the physical environment. Additionally, lighting is the strongest external cue for maintaining circadian rhythms, and hence future systems should provide appropriate light intensity and color temperature based on time of day, to promote optimal health. Spacecraft lighting systems should be designed to promote safety, task performance, and well-being. Meeting lighting needs within the space and power constraints of an exploration vehicle may involve some difficult tradeoffs. Mission objectives must be considered when addressing these tradeoffs.

Vibration and G-forces

When vibration is at an intensity and duration sufficient to impair visual perception, manual performance, or speech intelligibility, it will impede the conduct of normal duties. If severe enough, vibration may directly cause injury via mechanical stress to the internal organs and the musculoskeletal structure, or because limb flail results in impact with cabin equipment and adjacent crewmembers. Crewmembers experience significant vibration during launch, launch abort, and entry-phases of flight that are also accompanied by elevated G-loading. During these dynamic periods of flight the crew must interact with vehicle systems and provide prompt and accurate responses. Thus, the impacts on crew performance will depend not only on vibration and concurrent G-loading, but also on the design of the launch and entry vehicle, as well as the design of crew interfaces and required crew operations.

During launch and launch abort for all vehicle designs and during Earth entry in capsule type crew vehicles, the crewmembers lie in a semi-supine (recumbent) posture and experience hypergravity in the *x*-axis (i.e., chest-to-spine direction). These dynamic periods of flight can be accompanied by substantial vibration, potentially in all body directions, from propulsion systems and aeroacoustics (i.e., the aerodynamic interactions between the atmosphere and the surfaces of the launch and crew vehicle), and the structural response of vehicle systems to these vibration inputs. NASA has expended significant engineering effort on limiting vibration of the propulsion system (i.e., the pogo oscillation in liquid-fueled rocket motors for *all* crewed spaceflight programs beginning with Gemini and thrust oscillation in solid-rocket motors for future programs) to protect both the vehicle and the crew. Studies have also been conducted to understand the consequences of narrowly tuned (i.e., single frequency), uniaxial vibration on several aspects of crew performance and health.

Noise interference

Noise interference can affect performance of tasks that require communication, can affect cognitive functioning, and can contribute to long-term health risks such as hearing loss. Loud noises can cause temporary threshold shifts. Noise can also impact the crew's ability to access information, and has been attributed to loss of sleep (Barger et al. 2014).

Seating, Restraints and Personal Equipment

When the design of seating, restraints, and personal equipment do not accommodate the astronaut, this can prevent effective and efficient performance inside or outside the vehicle and create unsafe situations. Crewmembers experience long periods of recumbent and restrained static loading while seated during certain mission phases that could pose ergonomic risks of discomfort, performance decrement, and injury. Restraints may be lacking, poorly designed, or too complex and time-consuming to set up and use, causing tasks performed in microgravity to be more difficult and frustrating than necessary. If equipment such as a portable breathing apparatus or a sleep station is designed without other relevant hardware components or the human operator in mind, it can create ergonomic accommodation issues during integration and operational phases. Long-duration use of seating, restraints, and other equipment can cause the crewmembers to develop repetitive strain injuries (RSI).

Visibility/window design & placement

If the lighting system, windshield/window, glare, reflection, or other visual obstructions prevent necessary visibility, this can contribute to error, injury, or poor task performance. Critical visibility conditions, once identified through concept of operations development and task analyses, may be modeled with enough fidelity to assess the lighting and visibility necessary for task completion.

Vehicle/habitat volume/layout

When the design and layout prevent effective and efficient performance of crew tasks outside or outside the vehicle, this could create an unsafe situation. Beaubien and Baker (2002) cite inadequate volume in which to live and work as a stressor contributing to crew discomfort. Inefficiencies in spaceflight vehicle and habitat architectural design can affect crew safety, efficiency, and habitability.

Crew accommodations vary based on many factors, including the specific task design, whether the tasks are completed in a co-located areas, and the physical characteristics of the person. Appropriate habitat configuration and adequate provision of volume, whether inside or outside of the vehicle, is imperative to ensure compatibility with the characteristics and capabilities of the crewmembers and the necessary tasks they will perform. Insufficient NHV and inappropriate functional arrangements can affect productivity and habitability. Tasks that are unique for LDEMs must be considered in vehicle habitat design.

Evidence presented in the following section encompasses lessons learned from 50 years of spaceflight experience and ground-based research related to the risk of incompatible vehicle habitat design. Evidence consist of summaries of subjective experience (opinion), and non-experimental observations or comparative, correlation, and case or case-series studies (observational). Some evidence in this chapter is derived from the FCI ISS crew comments database (opinion); although summaries of ISS crew feedback are presented as evidence, the database is protected and not publicly available due to the sensitive nature of the raw crew data it contains. Data is also presented from Crew Office-approved Space Shuttle crew reports. These reports are not publicly available.

Much of the spaceflight data is opinion data acquired from the FCI ISS crew comments database, with some additional data based on spaceflight experiments. Spaceflight evidence is garnered from Space Shuttle and ISS flights, none of which to date exceed 6 months in duration. Thus, spaceflight evidence must be considered within the framework of future research needs for concepts such as an Evolvable Mars Campaign, which includes longer duration data and insight into additional environments such as partial gravity habitats.

Risk of Inadequate HSIA and Human Computer Interaction (HCI) Factors

HSIA must support autonomous operations during LDEMs, and information must be available on board that was previously provided by MCC. The communication delays expected on LDEMs will likely mean the crew will depend much more on computer-provided information. Crew may have to rely solely on available electronic information for just-in-time training, task procedures, and maintenance. The "safety net" of calling ground control for questions, workarounds, and forgotten procedural steps may no longer be feasible. That information will need to be well-designed and tailored for crew use.

HCI encompasses the processes involved in how humans and computer-based systems communicate, share information, and accomplish tasks. Information Architecture (IA) is the categorization of information into a coherent, intuitive, and usable structure. When HCI or IA is poorly designed, crewmembers have difficulty inputting, navigating, accessing, and understanding information. Information is presented most effectively when the users' interests, needs, and knowledge are considered in the interface design. If information displays are not designed with a fully developed operational concept, fine-grained task analysis, and knowledge of human information processing capabilities and limitations, the format, mode, and layout of the information may not optimally support task performance. This could result in users misinterpreting, overlooking, or ignoring the original intent of the information, leading to increased task completion times and necessitating costly replanning and rescheduling. This could also lead to errors when executing a task that may jeopardize crewmember safety and mission success.

Although much is known about designing systems that provide adequate HCI, LDEMs such as Artemis missions bring new challenges and risks. Whereas the space shuttle had hundreds of hard switches and buttons, exploration vehicles will have mostly glass-based interfaces, requiring the crewmember to rely on an input device to interact with software displays and controls (Ezer, 2011). Due to restrictions in mass for LDEMs, the real estate for displayed information is likely to be limited, but the amount of information available for display will be greatly increased, posing challenges for information design and navigation schemes. Furthermore, users will perform semi-automated electronic procedures that have not yet been fully proven in spaceflight. Future vehicles will also fly many new technologies that must be usable while wearing pressurized gloves, in microgravity, and during vibration.

Inadequate HCI can have a wide range of undesirable consequences. A significant risk of errors or failure of mission objectives will occur when the crewmembers cannot perform a task because they have trouble seeing or hearing needed information, when wrong information is displayed, when data is unavailable, or when the presented information is confusing. These errors may occur due to inadequate IA that leads to poor organization of information, or could relate to inadequate presentation of information. Risks can also be associated with the design of human interfaces if crewmembers cannot reach controls, have difficulty manipulating them, or if controls have unexpected behaviors, are poorly labeled or confusing, or are not available when needed. Additional problems arise when there is improper function allocation between the human and the system, or when the means of interaction with the system is confusing, inefficient, or difficult to learn or remember. These problems are exacerbated when procedures are poor, timelines are challenging, environments are unpredictable or dynamic (e.g., lighting, vibration), or crewmembers are stressed due to unexpected events.

Poor HCI can also reduce efficiency and undermine the added value of computer functionality if extraneous or overhead tasks are imposed on the user, requiring them to focus cognitive resources on something other than the task at hand (i.e., navigating or managing the user interface, or reorganizing information before proceeding with the task). Extraneous tasks can also be imposed if the user must integrate information from multiple sources or the interface requires significant attentional resources. Unfortunately, these sources of overhead tasks are not easy to detect, and often remain uncontrolled, posing a risk to users and to mission objectives. HCI task overhead negatively affects usability, therefore designers should try to minimize this occurrence whenever possible (Butler et al., 2007).

Consequences of inadequate HCI become far more serious during dynamic phases of flight such as launch, docking, and landing, when little time is available for correcting mistakes. As mission length increases, and ground support decreases, a comprehensive HCI approach needs to be captured in the HSIA because ground-assisted workarounds may not be possible, and mission objectives and crewmember safety may be impacted.

When designing LDEMs systems that support autonomous operations, several human-system areas need to be considered: (1) Informational resources and support, (2) Attention and cognition, (3) Motor skill, coordination and timing, (4) Environmentally induced perceptual changes, and (5) Design of displays and controls. These factors can prevent the user from successfully accomplishing tasks or task objectives because they affect the user's ability to properly use information to make correct decisions. A general description of each of these contributing factors follows, with examples specific to flight and ground instances presented in the next chapter.

Informational Resources and Support

Adequate informational resources and support becomes important when task information, operational planning material, or other information necessary for safe operations is not available. This factor is in play when the user does not have the information needed to perform a task because the information cannot be observed, is not provided, is not understandable, or is incorrect, or when crew-to-crew or crew-to-ground communication issues arise. Informational resources include visual and auditory information, procedures, schematics, and crew communication. These informational resources must be provided when the task cannot be completed without them, i.e, without the required information the user risks inducing errors or cannot complete the task.

A critical information resource is the 'procedure'. Whether paper or electronic, procedures drive virtually all major operations in a vehicle cockpit. They can be developed to aid performance if they include functions that track the current step, provide reminders in multi-tasking situations, provide information for decision making, automate certain processes, etc. Automated procedures must be considered carefully, because, if not implemented correctly, they could lead to increased workload, decreased efficiency, and increased risk. Electronic procedures for use in LDEMs will have to be carefully developed and extensively researched to ensure the safety of the crewmembers and efficient task completion.

When information is presented in the wrong format and excessive time and effort is needed to derive the meaning, there is increased risk of error and inefficiency. For example, if the crewmember needs to know velocity but only position and time are displayed, the crewmember has to do mental calculation to estimate the rate of change; if the temperature is digitally displayed and the crewmember needs temperature trends, the crewmember does not have the information needed. Computers are much better suited for these types of mathematical operations, whereas humans are better suited for more complex analysis and decision making (Kantowitz & Sorkin, 1983). Functions should be allocated accordingly.

The format for presenting information is particularly important in emergency situations when alerts are used. Traditionally, alerts have been presented in the form of lights and tones. Newer technologies make it possible to present alerts in other modalities. For example, speech alarms and tactile modes of presenting information can augment the visual information available to the crewmember. Speech alarms can specify the alert type along with the location or safety information. Tactile displays can alert and direct crewmembers when other modes of communication are not possible. These modalities can be combined into advanced alert systems that will be increasingly important during LDEMs where crewmembers must operate with greater levels of autonomy in various extreme environmental conditions such as low visibility or high noise. Research-based design guidelines must be developed for these advanced multimodal information systems.

Attention and Cognition

A crewmember's attention may be diverted if they focus on some cues in the environment and ignore other cues of equal or higher importance. This channelization of attention, called cognitive tunneling or attentional tunneling, can lead to an unsafe situation in which the individual is unable to develop comprehensive awareness of the situation and respond appropriately to critical events. This lack of alertness or readiness to immediately process available information could be due to a false sense of security. Research on heads-up displays, 3-dimensional displays, and fault management, for example, demonstrated that users may focus attention on some aspects of the task (e.g., visually compelling display elements) to the detriment or complete obliviousness to other aspects or events (Wickens, 2005). In other situations, the nature of the task or reliance on automation may cause the individual to fail to redirect their attention, recognize an automation failure, or seek additional information that could improve decision making (Sarter at el., 1997).

Anticipating problems that may occur due to improper allocation of attention during LDEMs is challenging. Whereas ISS crewmembers' time is generally overbooked because so much crew time is required to maintain the ISS, a LDEM may have a very different operational tempo. There will be dynamic phases of flight that require focused attention, and then long periods of time during the transit when crewmembers will have few attention-demanding tasks. During transit, crewmembers must still be alert and able focus their attention on unexpected emergency events.

A human's physical, sensory, perceptual, and cognitive capabilities have constraints that are associated with performance deficits. Regarding cognitive capabilities, for example, the amount of information that can be processed is limited by working memory (Baddeley, 1992; Miller, 1956), and information overload can occur when trying to accomplish tasks that load the working memory. In addition, operators can experience cognitive overload when they need to process a lot of information in a short time. The amount of information that an individual is able to acquire and process may be affected by stress, fatigue, time constraints, and the modality of that information (e.g., visual or auditory). In extreme cases, confusion can occur when the individual is unable to maintain a cohesive and orderly awareness of events and required actions and experiences; a state characterized by bewilderment, lack of clear thinking, or disorientation. Lack of transparency or predictability of a given task can also decrease the users' understanding of the state of the system, contributing to task overload and confusion (Dix et al., 2004). During periods of confusion, an individual's performance on one or multiple tasks may be considerably reduced (Wickens, 1991).

In addition to cognitive overload, cognitive underload may also be a concern for crewmembers on a long-duration space transit (Gore, et al., 2018). Unlike ISS crewmembers, LDEM crewmembers may not have a steady flow of maintenance or other tasks to be performed. They may become bored, resulting in stress, fatigue, and reduced alertness, and their readiness to perform or respond to an emergency situation may be compromised. Cognitive underload can lead to decreased vigilance and can lead to loss of SA, i.e., being less aware of important aspects of the environment needed for the current task and future actions (de Winter et al., 2014).

We need to determine the "optimal" level of cognitive load for LDEM crewmembers to ensure that future spacecraft systems are designed accordingly. This will be particularly important due the autonomous nature of the LDEMs, when crewmembers, after a long uneventful transit, may have to perform complex duties under stress, fatigue, potential deconditioning, and time constraints. Human capabilities and limitations can be affected greatly by the duration of a mission and the degree of subsequent deconditioning. The strength and aerobic power of the crewmembers' load-bearing muscles can decrease during spaceflight missions. In-flight exercise regimens have been implemented to counteract these deficits, but to date have been only partially effective. Overall, the long-term effects of living in space and the effects on performance are still generally unknown. We do know that perception in every modality, reaction time, motor skills, and workload can be affected while in space, and thus can affect performance (Legner, 2004). It is important to understand how tasks, procedures, and schedules may need to be modified as deconditioning occurs.

Motor Skills, Coordination and Timing

Decrements in motor skills, motor coordination, and timing occurs in microgravity (Lackner & DiZio, 2000). The laws of motion of the body and of objects change in microgravity, creating distortions of orientation and posture, as well as disruptions of certain aspects of limb proprioception and oculomotor control (Lackner & DiZio, 2000). Furthermore, the ability to intercept or avoid a moving object is impaired in microgravity because the ability to anticipate the trajectory of a moving object is based on sensorimotor functions that were developed in a 1-G environment (Rushton & Wann, 1999). Motor skills will become increasingly important as microgravity exposure increases with longer duration missions. In addition, as crewmembers become increasingly deconditioned, detriments to their motor skills are possible that could affect interaction with displays and controls as well as with other system components. Deconditioned crewmembers may not have the fine motor skills required to perform critical tasks in a timely fashion. Whenever possible, these possible decrements in crewmembers' capabilities need to be considered when LDEM systems are being designed. To do this, we must have a good understanding of how long-duration spaceflight affects motor skills and of the mitigation strategies that might be appropriate.

Microgravity affects physical movement during spaceflight; it slows movement time (Berger et al., 1997), decreases tracking performance (Manzey et al., 1998), and increases variability (Semjen et al., 1998). The fine motor skills that are necessary for precise interaction with systems may be affected in microgravity. These fine motor skills will be especially important during LDEMs, during which crewmembers must be self-sufficient, and interact with systems such as spacesuits, space vehicles, and habitats. Holden et al. (2018a) determined that gravitational transition induced significant detriments to ISS crewmember's fine motor skills, and Moore et al. (2019) saw similar motor performance decrements in crewmembers at landing after return from 6-month missions. Clearly, exposure to microgravity and gravitational transition affects motor performance. This needs to be studied in more detail so that CMs can be developed for use during LDEMs.

Environmentally Induced Perceptual Changes

Stimuli in a microgravity environment can cause an erroneous perception of orientation, motion, or acceleration. When crewmembers experience altered perceptions of orientation, motion, or acceleration, their ability to interact with systems may be affected, namely the ability to read displays and to manipulate controls. Although NASA Ames Research Center (ARC) and NASA JSC have conducted a few studies on how whole-body vibration affects cognitive and manual performance (described in sections that follow), we still don't know how gravitational transitions during different phases of spaceflight affect short-term and long-term performance in various domains such as readability, perception, judgment, and decision-making ability.

Design of Displays and Controls

Design of displays and controls consists of the visual, auditory, or tactile aspects of the design. Because displays and controls are the primary interfaces through which crewmembers control vehicles and habitats, adequate design of these components is critical. When a human-centered design (HCD) process is not used during development, the design of displays and controls can be inconsistent, ineffective, inefficient, and unacceptable to crewmembers.

If displays and controls are inadequately designed, inadequate external representation can lead to a misunderstanding of a given scenario and may affect the user's problem-solving capability. Zhang and Norman (1994) defined distributed cognitive tasks as those that require processing of information that is distributed across the mind and the external environment. These tasks require both internal and external representations. Internal representations include propositions, productions, schemas, mental models; external representations include physical symbols (e.g. letters, numbers) or external rules, constraints or relations embedded in physical configurations (e.g., spatial relations of written digits, visual and spatial layout of diagrams). In their paper, Zhang and Norman discuss the representational structure of the Tower of Hanoi problem and show that representing the problem in a different way can dramatically affect the difficulty of solving the problem, even if the formal structures doesn't change. Similarly, the design of displays and controls can dramatically influence task difficulty.

Proper design of displays and controls includes ensuring that designs meet applicable standards and guidelines, were developed as part of an HCD process, and were verified with human performance measurement methods that are rigorous but feasible in an operational test environment. Human performance measures such as usability, legibility, workload, and errors may be simple to collect in a laboratory but can be more challenging to obtain in an operational test environment.

We know how to design adequate displays and controls for terrestrial applications, but space applications require special consideration, especially when the systems being designed will be used by crew to integrate information from multiple sources, control multiple systems, make decisions, and resolve unexpected anomalies. The design of displays and controls can ultimately mitigate the effects of all the design considerations described in the sections above. Displays must provide all the information needed for the crews to operate autonomously and provide the information in a form that is intuitive and promotes proper attention and cognitive load. Displays must also be usable in the variety of environmental conditions expected during LDEMs. Controls must be designed to provide automation when needed to compensate for the effects of the space environment. Research is needed to determine how displays and controls can be designed and must be verified to serve this complex mitigation role.

Risk of Inadequate HSIA and Human-Automation and Robotics Factors

As space exploration reaches more distant destinations, growth in mission complexity will demand flight crews and systems to be more autonomous. Increased autonomy will necessitate heightened reliance on automation and robotic systems for both crewed and un-crewed missions. For space exploration, automation increases productivity and reduces human workload by supporting a broad range of functions and operations, from controlling flight vehicle subsystems (e.g., propulsion, environment control, etc.) to conducting science experiments (e.g., remote sensing, data processing). As a special class of computer-controlled machines, robots are an extension of automation. Robots expand human capabilities by extending mobility, enhancing manipulative strength and dexterity, and enabling ground and astronaut operators to act at a distance from the worksite, improving mission productivity.

The success of future LDEMs will depend on effective integration of humans and automation and robotics technology. This will not emerge by chance, but by design. Robotic systems and their user interfaces need to support multi-agent team configurations, widely separated operators and robots, varying communication delays, and variable levels of automation (LOA). Similarly, the integration of automated systems and their human users must allow for dynamic allocation of authority and autonomy to suit specific tasks and context.

During current NASA missions, the crew in space and controllers on the ground interact with automated and robotic systems to accomplish mission goals. Future missions LDEMs will require increased dependency on automation and robotics. NASA's new "Moon to Mars" initiative requires innovative and sustainable approaches to Earth-independence (https://www.nasa.gov/topics/moon-to-mars). Hence, future NASA operations involving crew-robot and crew-automation interactions will increase substantially relative to current operations (Marquez et al., 2017).

Human-Automation Integration (HAI) Design

Automation is the use of machines or computers, generally for the purpose of increasing productivity and reducing human cognitive workload. Automation typically includes a preplanned set of instructions, such as those used to operate a spacecraft or control deep-space robotic missions. NASA defines autonomy as "ability of a system to achieve goals while operating independently of external control" (Fong, 2018). Although AI is the computer science field dealing with simulation of intelligent behavior of machines or computers, many of its methods (for example, machine learning, object recognition, probabilistic methods, and neural network searches) are leveraged in both automation and autonomous systems. This risk of inadequate HSIA report focuses on current evidence regarding human-automation and (by extension) robotic integration, while projecting the risk of future human-autonomy teaming.

Current spaceflight crews rely on onboard automated systems. As increasing numbers of automated systems are designed to assist the human, a synergistic relationship must be developed between the human and automation so the two can work together to successfully accomplish tasks. On future missions that are longer and more autonomous, crews will rely even more on these systems to provide information that is appropriate, accurate, and up to date. Wickens et al. (2010) found that “routine” performance was improved, workload was reduced, and SA was increased by introducing automation. However, if all tasks are automated, humans can become complacent and lose SA (Parasuraman et al., 1993; Parasuraman & Riley, 1997). In addition, increased automation will require special emphasis on task design and additional training to ensure that the crew can perform the automated tasks if the automation fails. Automated tasks must be carefully designed to prevent the crew from losing SA or becoming unaware of, or complacent about, potential hazards—situations that could ultimately result in system errors, degraded crew performance, and compromised crew and vehicle safety. A specific requirement for future LDEMs is automated planning capabilities and tools. These tools will help autonomous crews by providing alternative plans and solutions for managing daily tasks.

Different types of automation will be required to support the highly diverse set of functions and operations during LDEMs. The required automation systems will span ground and flight systems and support functions ranging from controlling the habitat to conducting science experiments. The integration of automated systems with their human users will necessitate a variety of role divisions: the allocation of authority and autonomy between human and automation may change dynamically depending on task or the context. Increasing automation within a spacecraft does not mean that human error will be eliminated or that people will no longer be needed. Inherently, automation shifts the nature of the work people must conduct, a critical trade that is often overlooked. Failure to address the HAI will lead to ineffective and inefficient systems.

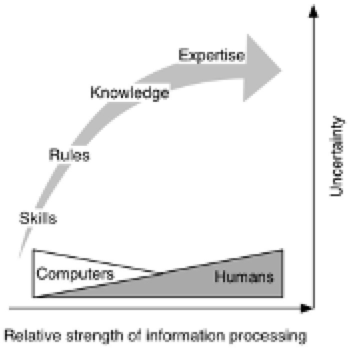


Figure 3. Relative Strengths of Computer and Human Information Processing, Adapted from Cummings, 2014, 2019a

When investigating human performance with respect to automation, the risks associated with autonomous systems are often downplayed (Cummings, 2017; Woods, 2016; Woods & Dekker, 2000). Woods points to “Doyle’s Catch” (Alderson & Doyle, 2010) and summarizes this quite succinctly: it is "*presumed* that because [autonomous] capabilities could be demonstrated under some conditions, extending the prototypes to handle the full range of complexities that emerge and change over life cycles would be straightforward" (Woods, 2016, p. 131). There is a gulf of uncertainty within the life cycle of one autonomous system, and furthermore, the advances of fruitful implementations of automation in one domain will not necessarily translate into success in spaceflight operations. It should be assumed that automation and AI are only as good as their implementation. Moreover, there is risk in assuming that AI is "all powerful" (Cummings, 2019b). As uncertainty increases, automation is less effective and humans are more effective because they are adaptive and flexible, as illustrated in 3. Woods (2016) points that multiple organizations, both in government and industry, fail to properly consider automation brittleness and its contribution to human-system integration.

Furthermore, AI and machine learning, the basis for much of the forthcoming autonomous systems, require a large amount of training data to develop their reasoning algorithms. Insufficient training data leads to less reliable results. Additionally, it has recently been recognized that the resulting algorithms are biased because the training data used is systematically biased (Knight, 2017a). This introduces safety concerns as AI spreads to critical areas like medicine (Knight, 2017b). Inevitably, when these autonomous systems encounter an anomalous condition or the conditions and the context of use change, it is the human operator who has to jump in, adapt to the circumstance, identify the fault, and resolve the problem.

Human-Robotic Integration (HRI) Design

Robotics is an extension of automation. Hancock et al. (2011) characterize robots as mobile devices, having a human or animal form, that are intended to "effect action at a distance;" the latter attribute being of significant importance for space and planetary exploration. Dexterous, heavy-lift, mobility systems are the classes of robotic systems expected for near- and deep-space exploration missions. Human-robot teaming will be a cornerstone of future operations. Robotic systems and their interfaces must be designed to support all levels of human operation (direct manual control, shared control, and supervisory control). These systems must also support multiple robot operators in multi-agent team configurations, operators physically distributed near and far, varying time delays, and increased reliance on high and variable LOA.

An appropriate analogy to successfully designed HARI is the "H-Metaphor" (Flemisch et al., 2003). This analogy is derived from considering a horse as an autonomous vehicle; a person may guide and direct the horse, yet the horse will provide feedback to the rider, preventing dangerous situations. The H-Metaphor can remind designers to consider the operator and the intelligent system as cooperative agents that share the same goal. Similarly, complex automation and robotic systems should be designed to enable the operator to be part of the system, providing guidance and receiving feedback.

Ineffective user interfaces, poor system designs, or ill-advised functional task allocation compromise successful human performance, and subsequently, mission success and safety. As a result, the design of effective human automation and robotic design is essential to ensure humans can perform nominally and under the deconditioned cognitive and physical states expected during LDEMs. Gaps in our knowledge and experience exist for the expected level of complexity of future automated and robotic operations (Karasinski et al., 2019; Marquez et al., 2017). Concerning robotics, risk arises because we have limited experience with teleoperation beyond robotic arms, time-delayed operations, and multi-robot operations. For example, poorly designed human interfaces can result in a loss of SA, compromising mission safety and efficiency. Of special concern are losses of SA while a crewmember is near to a robot. Which could risk the crewmember’s safety. Crewmembers must be able to ascertain and understand the state of a robot, affect or change its state, and override the system whenever necessary.

In anticipation of crews operating more autonomously during future LDEMs, HFBP has started the HCAAM project (Holden et al., 2019a; Holden et al., 2019b). In 2019, 7 research projects were awarded to assess how automation and advanced robotic systems affect human performance in spaceflight settings. Mission complexity will increase as mission familiarity and experience decrease, introducing increased challenges to effective design of automation and robotics and increased the risk of inadequate human performance in spaceflight. A critical aspect to meeting automation design challenges is a focus on human-system integration (Terwiesch & Ganz, 2009). The systems, automation, and robotics, must be well integrated into the process and the workflow. Because complex missions will require multiple automation systems, robots, and people, it is critical to design with an awareness of how these multiple agents can and should be coordinated. Changing the roles and assignments lead to changes in the work, shifts times of peak human workload, and creates new, and often challenging human tasks such as supervisory responsibility. Hence, emphasis must be placed on integrating the design of these complex systems.

Risk of Inadequate HSIA and Inadequate Training Factors

Introduction to Spaceflight Training

During the Mercury, Gemini, and Apollo mission eras, crewmembers learned the vehicle systems as they worked collaboratively with system engineers on the vehicle design. In fact, the selection criteria for the Mercury 7 required astronaut candidates (ASCANs) to have an engineering background and test pilot experience (Weaver, 2015). Astronauts were assigned specific roles and provided numerous hours of training in high-fidelity, part-task trainers. They also trained as a team in full-mission trainers, practicing known mission timelines with the MCC, averaging almost 1,000 hours of simulator time per crewmember per mission during the Apollo Program (Woodling, 1973). The training required the crew to “practice and demonstrate proficiency in the critical aspects of the mission” (Woodling, 1973), and the simulators were provided to allow for the practice needed to achieve these training goals. The development of the Apollo simulation program and associated trainers closely paralleled the development of the Apollo flight hardware and the increasing mission complexity. As flight operations progressed from Gemini to Apollo, and as Apollo flights progressed from single- to dual-spacecraft operations, Earth-orbital to lunar-orbital activities, and finally to the lunar landing, the scope and capability of simulations matured to keep pace with the increasing mission complexity.

Apollo-era astronauts trained for rendezvous, docking, and lunar landing using both fixed-base and motion-base simulators. Motion-base simulators of rendezvous and docking included Langley Research Center's rendezvous and docking simulator, which consisted of a crew station supported in a gimbal suspended from a bridge crane (providing 6 degree of freedom [DOF]), and later the translation and docking simulator, which simulated 6 DOF of crew station and the Agena target vehicle. For lunar landing, a free-flying lunar landing training vehicle was designed to familiarize pilots with the handling characteristics of the lunar module during the final portion of descent and touchdown.

Although Woodling, et al. (1973) did not provide the complete rationale for using motion-base simulators, they did note: "Although the complexity and cost of these devices were significant, their use in mission training was mandatory" (p. 29). In the case of the lunar landing training vehicle, "danger" could be added to "complexity and cost." Woodling, et al. concluded that, "Although training for tasks such as reentry, docking, launch abort, and lunar landing has been accomplished on fixed-base simulators, crew proficiency for each of these tasks has been improved with moving-base simulators. The moving base devices were used to add realism to the simulation by presenting realistic cues inherent in body acceleration and binocular vision" (p. 30).

"Practice, practice, practice" was the unofficial motto of Apollo-era training (Weaver et al., 2015). During this time, NASA made tremendous engineering advancements in designing simulators that could mimic the various spaceflight environments the crew would encounter (e.g., neutral buoyancy simulators EVAs (aka space walks), motion-based simulators for ascent/entry (described above), and reduced gravity simulators for lunar surface operations) (Weaver, 2015; Woodling et al., 1973) allowing the astronauts to practice for their well-planned and choreographed missions. Missions were accomplished with the onboard crew working as a team with the experts in the MCC.

During the Space Shuttle program, newly hired ASCAN were provided 2 years of ASCAN training prior to being designated as astronauts and being qualified for assignment to a Space Shuttle mission. The first Space Shuttle crews learned the vehicle systems as the vehicle was being designed, but as the Space Shuttle program progressed, spaceflight training was formalized into sets of lessons designed into system or discipline training flows (e.g., an electrical power training flow, an avionics training flow, a robotics training flow, and an EVA training flow), and training culminated in a series of full team simulations that provided repeated rehearsal (or “hammering in”) of their specific roles for well-planned and detailed mission timelines.

Flight-specific training for a Space Shuttle mission typically lasted one year, although flight delays would extend this training time. Crewmembers were provided numerous hours of training to practice their assigned individual duties and team tasks with MCC. Training simulations where provided to practice ascent and entry timelines as well as challenging longer-timeline sequences such as satellite deployments or EVA per specific flight plans for missions that lasted up to 17 days.

Astronauts were provided training in part-task trainers, mockups, and simulators of varying fidelity, including those designed to mimic the spaceflight environment (e.g., microgravity via parabolic flight, motion-base simulators) and based on engineering advancements made in earlier programs (National Research Council [NRC], 2011; Woodling et al., 1973). Astronauts practiced approach and landing piloting skills using a suite of training simulators: a fixed-based simulator called the Shuttle engineering simulator; 2 motion-based simulators called the Shuttle mission simulator (SMS) and the motion-base vertical motion simulator; the flight-based shuttle training aircraft simulator; and T-38 aircraft.

The tremendous successes of the Apollo-era and the Space Shuttle programs proved that this training design in which crewmembers practiced known mission timelines as a team with MCC in simulators of varying fidelity was very effective for short-duration, near-Earth, and lunar missions.

For ISS missions, astronaut training again begins with a 2-year ASCAN training flow after which astronauts are qualified for flight assignments (Dempsey & Barshi, 2020). Although flight assignments can occur immediately after ASCAN training is completed, more typically an astronaut will be assigned office duties and continue training for up to 6 to 8 years (or more) before being assigned to a flight. Pre-assignment training during this period includes refresher training on ASCAN skills (e.g., ASCANs assigned as capsule communicators [CapComs] receive refresher training on ISS emergency response), training as part of assignments (including CapCom certification), and training in analogs such as the NASA Extreme Environment Mission Operations project and the Cooperative Adventure for Valuing and Exercising human behavior and performance skills course (Landon et al., 2018). Flight-specific training for ISS missions takes approximately 2 years. Flight-assigned training at JSC covers a wide range of systems, medical operations, robotics, EVA, and payloads, and also includes T-38 jets, mission support, language, and physical fitness training, and NASA trains crews in a large number of fixed-base simulators, microgravity simulators, and virtual reality (VR) simulators, as well as providing piloting training in T-38s (see Table 5).

Table 5: NASA's ISS Training Facilities and Simulators at JSC

	T-38	Neutral Buoyance Lab	Space Vehicle Mockup Facility	Space Station Training Facility	Virtual Reality Lab	Dynamics Skills Trainer, ROBoT, Robotics Flight Controller Trainer, Dome	Sim City	KBRwyle Medical Lab	Exercise Lab	Flight Controller Part Task Training	Ops LAN Part Task Trainer	Part Task Trainers
Piloting	X											
Medical Operations								X				
Countermeasures (Exercise)			X						X			
EVA		X	X	X	X					X		
Robotics		X		X	X	X	X			X		
Maintenance and Repair			X	X						X		X
Vehicle Systems			X	X			X			X		X
Operational LAN			X	X							X	
Inventory and Stowage			X	X							X	
Crew Systems and Photo/TV			X	X								X
Visiting Vehicle Operations							X					

Since early in the ISS program, ISS astronauts and cosmonauts have flown to the ISS on the Russian Soyuz vehicle. Although ISS docking operations are nominally automated, Russia trains cosmonauts and astronauts for contingency manual-control operations in a ground fixed-base simulator and a ground centrifuge simulator, where the only controllable parameter in the centrifuge is *g*-load (Wenzel, 2019). Additionally, Russia provides an onboard simulator for proficiency training.

In May of 2020, NASA astronauts again began piloting spacecraft launched from American soil on Space-X's commercial crew program Crew Dragon spacecraft. Astronauts will also be piloting Boeing's commercial crew program Starliner spacecraft and NASA's Orion spacecraft. NASA, Boeing, and Space-X are currently providing ground fixed-base training simulators for astronaut training. NASA is also planning the design of astronaut training for NASA's Artemis program that will return humans to the lunar surface.

The ISS training program continues to be structured based on system or discipline flows developed in the Space Shuttle era, and these training flows are coupled with practice or rehearsal in training simulators of varying fidelities. Certified training instructors develop flows based on a well-established instructional system design model, and certified chief training officers lead integrated training simulations in low-, medium-, and high-fidelity mockups that provide for multi-disciplinary training. Training is highly dependent on mission parameters, and ISS missions differ significantly from NASA's previous missions in several important ways that impact the design of training. Most notably, the ISS mission are significantly longer (typically ~6 months) and, for operational flexibility, astronauts are not assigned unique roles or duties, and ISS operations are ongoing— detailed flight plans, or timelines, are developed real-time based on long-range plans. NASA's spaceflight training team has shifted from "hammering in" known, short-duration mission timelines, to providing training on anticipated skills that are based on long-range plans and training every crewmember on a much wider range of skills, including EVA, robotics, scientific research, repair and maintenance, and potential malfunction and emergency response (see e.g., NRC, 2011).

Although the ISS crew training program produces highly skilled and effective crewmembers, instructors in NASA's Flight Operations Directorate (FOD) recognize issues with the crew retaining the extensive amount of content of the ISS preflight training, and that real-time ground support from MCC is critical to the success of ISS missions (Dempsey et al., 2018). With this real-time support continually available, both as a countermeasure against these training retention issues and to provide in-depth system expertise in the event of unanticipated failures, the ISS program of training is very effective for Earth-reliant operations.

Decades of success in human spaceflight have shown that NASA's spaceflight training programs are effective for low-Earth orbit and lunar missions. However, to date, the ground team, typically consisting of 40–50 flight controllers for the ISS, has always been able to provide real-time expertise to the onboard crew as needed, and a return to Earth has always been an option in response to any unforeseen event. Mars, Earth's closest planetary neighbor, is up to ~24 light-minutes away, and at times completely blocked by the Sun. A quick return to Earth will not be possible for Mars missions, and real-time expertise from MCC will no longer be available.

NASA has never provided astronauts with the training necessary to operate at the level of autonomy that these future missions will require. As recognized by FOD, the current design of spaceflight training does not produce the level of retention necessary for future semi-autonomous to autonomous mission operations (see e.g., Dempsey et al., 2018). Most critically, the current program of training does not provide astronauts with the technical expertise and decision-making skills necessary to work outside published procedures and respond to unanticipated, high-risk, time-critical events without real-time ground support— events that the highly trained and skilled flight control team in MCC currently respond to (Dempsey et al, 2018; Vera, 2019; Wu & Vera, 2019).

Training Factors that Contribute to the Risk of Inadequate HSIA

Training is one of the key human factors areas identified as important to the HSIA risk (Vera, 2019). The HSIA risk outcomes are performance-related: there is a possibility of adverse performance outcomes including that crew are unable to adequately respond to unanticipated critical malfunctions or detect safety critical procedural errors. Given that performance in the operational environment depends on 3 criteria: the person, the training, and the environment (Baldwin & Ford, 1988; Ford & Weissbein, 1997; Blume et al., 2010), an effective spaceflight training program is a critical component to mitigating the HSIA risk. In this section of the report, we detail the training factors that contribute to the HSIA risk.

Although a training can be categorized in multiple ways, for the purposes of this report, factors that can either contribute to an effective training program or contribute to risk are grouped into 4 main categories:

- Foundational Principles of Training
- Training Program Design
- Performance Support
- Designing for Trainability

Each of these categories, or factors, and how they support effective in-mission performance are described next in further detail. If these factors are not properly understood, required, or properly employed, then there is a risk of adverse performance outcomes, as discussed in later sections of this report on spaceflight and ground-based evidence.

Foundational Principles of Training

Training retention, transfer of training, simulation fidelity, and expertise with automation, are foundational factors in the design of an effective spaceflight training program. Designers must have understanding of each of these factors and the tasks the crew needs to perform during their mission to design an effective program of training. Although these factors listed above are 4 separate issues, the measures of concern are retention of training and expertise as evidenced by in-mission performance.

Retention

Training can be defined as “the systematic acquisition of knowledge (i.e., what we need to know), skills (i.e., what we need to do), and attitudes (i.e., what we need to feel)” (Salas et al., 2006). The retention of training can be defined as the ability to recall the skills, knowledge, and attitudes learned in training. Training is intended to provide long-term retention of skills, knowledge, and attitudes; yet, even with an effective initial training design, long periods of disuse lead to skill decay (see, e.g., Winfred et al. 1998).

Declines in performance after learning, as exhibited by increased response time (or decreased accuracy), have been known since the time of Ebbinghaus (1885/1913), who used a measure of savings (i.e., the amount of relearning required to achieve the criterion level of performance during original learning). Subsequently, this relationship between response time and retention interval was described as a power law (Wickelgren, 1974), $R = (d + fT)^{-g}$, where R is response time, T is the retention interval, d is the criterion of original learning, f is a scaling parameter, and g is the rate of forgetting. This power law of forgetting (Wixted & Carpenter, 2007; see also Rubin & Wenzel, 1996) can be thought of as the inverse of the power law of practice (Newell & Rosenbloom, 1981), which describes the acquisition process for most skills (the relationship between trials of practice and time to make a correct response is a power function, $R = aN^{-b}$, where R is response time on trial N, a is response time on trial 1, and b is the rate of change). An understanding of the acquisition process and of the retention interval, especially for critical mission tasks, is crucial for the design of effective training.

Many personal characteristics can affect retention, including motivation, intelligence, and conscientiousness (see e.g., Ford et al., 2018; Hoffman et al., 2010). Astronauts are highly motivated, intelligent, and conscientious students. Nonetheless, even with such highly motivated students, memory retention must be supported for successful spaceflight missions.

Transfer

NASA's spaceflight training program is designed to ensure that astronauts can perform their assigned duties within the operational environment. However, as with most complex jobs, training does not, and cannot, cover all possible operational scenarios. The content included in preflight training assumes that any novel tasks encountered by the onboard crew are similar enough to the tasks covered in training and that the training transfers, or generalizes, to the novel in-mission task. Again, for ISS training, this assumes that real-time ground support is available as needed to support gaps in training retention or transfer.

Transfer and generalizability are two related concepts in training design that are often used interchangeably. Transfer of training most often describes the transfer of training from the learning environment to the operational environment; this concept is sometimes described interchangeably with training retention, although the learning environment is often not truly identical to the training environment and transfer from the learning environment requires more than just retention of a learned skill or task (Baldwin & Ford, 1988). Generalizability, most often describes the generalization of content learned in the training environment to novel tasks in the operational environment (Billman & Catrambone, 2019); this requires both retention of the trained material and generalization to the novel task. For the purposes of this report, we will use the term that most closely fits the cited research.

Training to acquire durable skills often leads to high specificity (see, e.g., Healy & Bourne, 1995; Vogel & Thompson, 1995), and these skills, although retained for extended times, do not generalize well beyond the specific training context. Understanding the relationship between retention and transfer, and how to support both effectively, is crucial for the design of effective training as described by Koedinger et al. (2012): "learning is robust when it lasts over time (long-term retention), [and] transfers to new situations that differ from the learning situation along various dimensions (e.g., material content, setting, cf., Barnett & Ceci, 2002; Chen & Klahr, 2008), …". (p. 761)

Simulation Fidelity

A simulation is the imitation of an environment, process, or event, whereas a simulator is the actual hardware and associated software that facilitates this imitation. Simulator and simulation fidelity are often used interchangeably when describing the level of realism that a training environment provides. However, it is useful to distinguish between them, because different levels of simulation and simulator fidelity might be required for effective training, depending on the skill or task to be trained.

Crewmembers could be trained on an isolated skill using a high-fidelity part-task simulator, e.g., training for spacecraft docking using an accurate representation of a vehicle's control system and visual and motion cues would provide low fidelity of the overall mission environment but high fidelity of the hardware and control software used for the actual task. Whereas, other types of training may not even require a simulator, e.g., symbolic rehearsal, a verbal or mental simulation of the task, could be used for training (Barshi, 2015), although this process would still require a high level of fidelity by going over every aspect of a procedure or task in detail. Despite these notions, simulator designers and operators generally still assume that the highest overall simulator fidelity leads to the best training. Simulating some aspects of an environment or a task with high fidelity might require high simulator fidelity. Likewise, aspects of the simulation (e.g., vehicle motion) often must be highly defined for the simulator (e.g., motion base) to produce high-fidelity cues.

Despite the considerable amount of research on ground-based simulators, not much is known about the simulation fidelity needs for the types of training and range of tasks associated with LDEMs. The majority of simulator fidelity research has focused on skill-based control tasks in aircraft simulations. In virtually all studies, simulator fidelity was considered without assessing the underlying relationship between the simulation (software) and the fidelity of the simulator (hardware and software). In addition, most research focused on how different levels of simulator fidelity affect human perception, behavior, and performance, and not in the context of training. Finally, the research to date focused on specific tasks and scenarios, not on generalizable skills.

Crews will have to practice high-risk, critical skills in simulators before LDEMs, and because these mission will be so long, onboard simulators will be required to refresh preflight training and to train crews during the flight on new tasks, including nominal, contingency, and off-nominal scenarios (Barshi & Dempsey, 2018). Given the limited resources, training programs most likely need to shift from only training very specific tasks to training more generalizable skills. Choices of levels of simulation fidelity across the training continuum (preflight initial, preflight refresher, onboard initial, onboard refresher, and just-in time training) must be based on thorough study and clear evidence of effective training. Furthermore, a clear understanding of the required simulator fidelity based on a given level of simulation fidelity needs to be acquired for the range of tasks associated with LDEMs. Incorporating team skills into training simulations will be essential to ensure effective team performance (Landon et al., 2018.)

Training for Expertise with Automation

It is important to determine the level of simulation fidelity necessary for LDEM training because many novel technologies and operational scenarios will exist for these missions. Extensive automated, intelligent, robotic, and decision-making systems may be used to assist crews in their work, and crewmembers will need training to develop expertise with this automation. Given that these technologies are still emerging, it is likely that the modeling and simulation platforms that can be leveraged for training on such systems will be limited or lacking.

Automated, intelligent, or robotic systems that are well-designed can still cause accidents if humans do not understand what the system is doing or how control is distributed between human and the system (Strauch, 2017). Even experienced, skilled, and motivated users could make an error if training does not provide these users a complete and accurate model of the automation, intelligent system, decision-making tools, and HAI, or if direct "hands-on" training with the system (or a simulation of the system) is not available. Training must align the user's mental model of the automated, intelligent system, decision-making tools, or robotic operations with how they are designed to function (see, e.g., Billings, 1997, Strauch, 2017). This includes teaching specific procedures, instilling in trainees an understanding of the environment and work to be done, and communicating a deep functional understanding of the strengths and weakness of automated and intelligent systems and of decision-making tools. Training should also include information about how control is distributed between the user and the automation, how the user may change control, what other factors influence the control state, and how to determine the current state of control.

Training Program Design

Training programs are designed to ensure that the trained personnel can perform their assigned duties in a given operational environment (see e.g., Baldwin & Ford, 1988; Salas et al., 2006). NASA's spaceflight training program for ISS missions is designed to ensure that astronauts can work together as an onboard team, and that the flight controllers in MCC provide real-time support to help the crew respond to high-risk, critical situations, to command, operate, and maintain their spacecraft, and to conduct scientific research in support of mission objectives. Kirkpatrick proposed "4 Levels" to evaluate a program of training (Kirkpatrick, 1959); the salient measures of concern for LDEMs are measures from the third level of Kirkpatrick's framework, measures of behavior on the job. Specifically, the measures of concern for the training program design are retention of training and expertise as evidenced by in-mission performance.

As described earlier, to achieve the high level of performance necessary for spaceflight missions, NASA astronauts are provided 2-years of training as ASCANs; once assigned to a mission, they are provided flight-assigned training across a wide range of disciplines for approximately 2 more years. NASA's training program is a well-established model that is designed based on the analysis, design, development, implementation, and evaluation of instructional systems (Salas et al., 2006). Three of these factors—the analysis, the design, and the evaluation of training—can affect the measures of concern for training program design, the retention of training, and the expertise; we describe these areas in more detail below.

Analysis: Duty and Task Allocation

Training needs analysis that can be used to define and develop the training content for a given program; analysis is conducted for a specific job described by a set of duties and tasks. In the Apollo and Space Shuttle eras, astronauts were given different jobs depending on their role in the mission. The duties and tasks of the commander were different from the duties and tasks of the flight engineer. For ISS astronauts, NASA shifted away from this concept, and astronauts were, and continue to be, trained for all duties and tasks, thus dramatically increasing the training burden on individual crewmembers. Given the difficulties of retaining all the training information for ISS missions, the expectation is that NASA will shift back to assigning astronauts different jobs with unique sets of duties and tasks, granting that some tasks exit that all crewmembers will need to be able to perform (Dempsey et al., 2018). If so, NASA will need a method or framework to determine the duties and tasks for each astronaut role.

Design: The Training Flow Across the Training Continuum

Once the training content is defined, the next step is to determine the flow of the training content. Training is distributed along a continuum, from initial preflight training, through ground and in-flight refreshers, to onboard initial training, to just-in-time training, and finally to performance support tools (Figure 4). Careful distribution of training topics and training methods is required across the full continuum of training. The design of the initial, preflight training relates to task-based training vs. skill-based training, effective use of simulation facilities, the methods by which the correct level of simulation fidelity can be selected given the operational context and the individual learner, and the correct training method and delivery system (e.g., classroom vs. computer-based training).

Figure 4. The Training Continuum from Barshi et al., (2017)

Training during spaceflight will be significantly constrained by the available onboard delivery technologies, although virtual environments may be implemented for this training. However, given the limited ground-based experience that has been gained thus far with such virtual training systems, key questions remain about training methods and delivery mechanisms, as well as about differences in acquisition and retention between on-ground and onboard training sessions. A systematic method is required to determine the topics that should be refreshed and at what interval refreshers are required. Just-in-time training requires the ability to expect the unexpected. Training can only be developed for expected tasks and situations; just-in-time training will be needed to deal with low likelihood events and situations that are not time critical. Because not all such events can be anticipated, methods must be designed so the crew can develop their own training to deal with these events when communication delays prevent information from being uploaded from the ground.

Somewhat similar issues exist for the development of performance support tools, either for situations that don't justify training, or for when time is critical (including emergencies) and the crew does not have the opportunity to receive training prior to operations. Again, determining which events can be handled using performance support tools and how such tools can best be designed requires systematic methodologies that do not yet exist.

Design: Training Methodologies

Current training methods for space missions are largely task or skill-based and rely heavily on practice or rehearsal. When training for ISS missions, crewmembers practice executing nominal procedures and responding to off-nominal contingencies. This rehearsal or practice design is how NASA trained crewmembers during the Apollo-era, and this process has changed very little in the intervening decades. However, numerous training methods, based on well-studied principles of learning, exist beyond rote practice and rehearsal, such as strategic use of knowledge, massed practice, spaced practice, and overlearning (see e.g., Hoffman et al., 2010; Kole et al., 2020). Three training methods—variability of practice, easy-to-difficult ordering, and focus of attention— are described below.

Variability of practice refers to a training design that provides varying conditions under which a task is practiced so that the task is not always practiced in the same manner (Kole et al., 2020). Although variability of practice slows the acquisition process, according to Kole et al. (2020) it is a "particularly powerful" training principle because it applies to both procedural and declarative memory tasks and supports both retention and transfer. Variable practice conditions include variations in the task itself, variations in the conditions under which the task is trained such as the sequencing of the task with other tasks, or even variations in similar but different tasks (Barshi, 2015; Healy & Bourne, 2012; Kole et al., 2020).

In varying the order of tasks during training, the easy-to-difficult ordering principle can be applied. In discussing the different views on the easy-to-difficult ordering principle, Healy and Bourne (2012) suggested, "Whether or not training should begin with the easiest or most difficult components of a fractionated task depends on task characteristics" (p. 20), and "trainers need to be sensitive to these characteristics before deciding on the order of the subtasks" (p. 20).

Some research suggests that novices should maintain an internal focus of attention by focusing on their body movements when first learning a motor skill; however, Kole et al. (2020) stated that once individuals have had at least some practice with learning a motor skill, they should maintain an external focus of attention when performing the task, focusing on the outcome of their actions rather than their body movements. This shift in focus of attention supports both retention and transfer of the trained tasks.

Training crewmembers using methods that support skill acquisition, retention, and transfer will be important for LDEMs.

Evaluation Methods

ISS crewmembers are evaluated in system-specific mastery lessons, i.e., they are evaluated not on their ability to perform the mission as a whole, but rather on each system or discipline, such as the electrical power system or the motion control system. They are also evaluated on their robotics skills and EVA skills. NASA does not currently provide a summative evaluation of individual crewmembers at the end of their training, or evaluations of teams of flight-assigned crews.

Training for other high-risk endeavors is evaluated throughout the program to determine the level of training achieved in the learning environment and the level of skill that transferred to the operational environment (see Kirkpatrick, 1959). If the effectiveness of the training measures, the methods of evaluation, or the timing of evaluations are not evaluated as suggested by Kirkpatrick for level 2 and for level 3, current evaluation criteria and practices may not be sufficient to meet the needs of new types of tasks, different mission operations, and changing training practices required for future missions.

Performance Support

The first 2 factors that contribute to an effective training program (foundational principles of training and training program design) stand alone as training concerns (Kirkpatrick, 1959). The next 2 factors that contribute to an effective training program (performance support and designing for trainability) impact training and are interwoven with other concerns across the HSIA risk. Performance support may be provided by other team members including the flight controllers in MCC or via onboard tools that include any information, device, or agent that supports task performance in the operational environment, including procedures, cognitive aids, information systems, and virtual assistants. The measure of effective performance-support is seen in in-mission task performance.

During complex operations such as EVAs, complicated maintenance tasks, and robotics operations on the ISS, MCC provides real-time ground support and continually monitors and communicates with the crew (Dempsey et al., 2018). Because real-time ground support will not be possible during LDEMs, enhanced performance support tools (including adaptable training and decision support tools) will be needed. The expectation is that such performance support tools will be readily available during LDEMs, and indeed research on cognitive aids and electronic procedures is well underway (e.g., Holden et al., 2018c). Although new technologies in augmented reality (AR) head mounted display (HMD) devices and intelligent systems offer promise to enhanced performance, performance support tools have not yet been developed for the capabilities needed to support autonomous mission operations.

The concept of onboard handover is a current paradigm of ISS operations that compensates for training deficiencies; however, this will not apply to LDEMs. New crewmembers arriving at the ISS are assigned time for formal handover during which the experienced crewmembers familiarize them with the vehicle and the general onboard operations. Additionally, the first time a new crewmember performs certain tasks such as using the exercise equipment, using the waste and hygiene compartment, or taking water out of a portable water dispenser, an experienced crewmember is often provides a functional handover of the task. Crewmember have commented that in certain instances they consider the assistance by an experienced crewmember to be “critical” to their performing the task correctly the first time (FCI ISS crew comments database, n.d.). Experienced crewmembers who can provide onboard assistance to novice astronauts will be unlikely during LDEMs. Thus, crewmembers will not have the benefit of such a handover and will have to be trained in advance to perform all such tasks.

Training is designed based on many factors including the performance support tools available to the crew (see, e.g., Baldwin & Ford, 1988). Effective performance support tools can reduce the training burden (i.e., reduce the overall content of training). Reducing the training burden can increase training retention, thus reduce the risk of adverse performance outcomes.

Designing for Trainability

The fourth factor that contributes to an effective training program is designing for trainability. Trainability addresses the extent to which the vehicle habitat, display interfaces, systems, and task designs support skill acquisition, retention, and how well they transfer to effective in-mission performance. The vehicle design should not be so overly complex as to be essentially untrainable given realistic constraints on training budgets and time. A vehicle that is effectively designed using HSI principles can reduce the preflight and in-flight training burden, which can increase training retention and thus reduce the risk of adverse performance outcomes.

NASA Policy Requirement (NPR) 7123.2, the NASA Systems Engineering Processes and Requirements, requires that projects have an HSI plan and defines HSI as,

> “An interdisciplinary and comprehensive management and technical process that focuses on the integration of human considerations into the system acquisition and development processes to enhance human system design, reduce life-cycle ownership cost, and optimize total system performance. Human system domain design activities associated with manpower, personnel, training, human factors engineering, safety, health, habitability, and survivability are considered concurrently and integrated with all other systems engineering design activities."

NASA Standards 3001, Volume 1, Revision A: Crew Health and Volume 2, Revision A: Human Factors, Habitability and Environmental Health (NASA 2015), are standards for human spaceflight missions directed at minimizing crew health and performance risks. There are over 600 distinct standards in NASA-STD-3001, more than half of which relate to crew training. Throughout NASA-STD-3001, standards or rationale for standards include references to reducing or minimizing the training burden, such as:

"**7.5.3 Medical Equipment Usability**
Medical equipment shall be usable by non-physician crewmembers in the event that a physician crewmember is not present or is the one who requires medical treatment.

Rationale: Medical equipment is to be simple and easy to use and require minimal training so that non-medical personnel can administer care to ill or injured crewmembers."

"**9.7.4.2 Failure Notification**
The system shall alert the crew when critical equipment has failed or is not operating within tolerance limits.

Rationale: An alerting system decreases the cognitive load on the crew: they do not have to try to surmise a system failure based on symptoms. Terminology, references, and graphics used are to be coordinated with other crew task demands so as to minimize additional training."

Thus, NASA NPRs and standards require that training be considered throughout the design of the project.

One example of an approach to design for trainability and to minimize the training burden is to reduce the design complexity. A successful example of this was demonstrated during a first of its kind space-ground experiment for robotic teleoperation: astronauts on the ISS commanded the German Aerospace Center and the European Space Agency humanoid service robot on the ground in various scenarios including complex surveillance and service and repair tasks in a simulated Martian solar farm (Schmaus et al., 2019). The astronauts received no previous training for these scenarios. Executing basic tasks and sequences autonomously, robots can significantly reduce an astronaut's work; however, commanding robots from Earth increases communication delays and overall workload. Researchers attribute this successful commanding from space to the simple intuitive interface design, which allows control of the remote robot with minimal cognitive load for the operator. The human-robot interface systems were designed with an object-centered rather than robot-centered intelligence of the robot. Incorporated into the design were standardized graphical user interface (GUI) elements (clear, visible, not cluttered), a filtered set of commands that were context-specific reasonable robot commands, augmented 3D objects to access object-related commands, and an intuitive GUI.

Summary of the Introduction of HSIA related Training Factors

Although NASA has a long history of successfully training astronauts for low-Earth orbit and lunar missions, NASA has never provided astronauts with the training necessary to operate at the level of autonomy that future LDEMs will require. To provide future LDEM crewmembers an effective program of training that develops the level of expertise needed to operate independently from the ground and that produces the level of retention needed for such long-duration missions, NASA must

- determine the retention interval for the skills trained to crewmembers of LDEMs, especially for mission critical tasks.
- provide simulators and simulations at the level of fidelity necessary for skill acquisition, retention, and transfer, including simulator training for expertise with automation.
- develop a method for allocating duties and tasks.
- ensure the design of training across the entire training continuum supports the level of retention and expertise needed for autonomous mission operations.
- ensure that methods supporting both retention and transfer are effectively employed.
- define evaluation measures and methods for critical mission tasks.
- provide effective performance support tools for in-flight task performance, including decision-making in response to unanticipated critical malfunctions.
- employ HSI design principles to support trainability of in-mission tasks.

Providing an effective program of training will reduce the likelihood of adverse performance outcomes resulting from a lack of training retention or lack expertise in the onboard crew.

Risk of Inadequate HSIA and Procedure and Task Design Factors

To provide adequate task design, we need to understand relevant human capabilities and limitations for performing tasks and how these may impact workload and degrade performance on LDEMs. We also need to understand the effect that other factors may have on human-system performance (e.g., automation, autonomous operations, reality-augmenting devices devices).

Mission design includes the high-level coordination, planning, and scheduling of spaceflight missions. Process design refers to the methods used to accomplish mission goals and objectives. Task design refers to the specific steps used to accomplish a discrete task of interest. Tasks of interest include those that are necessary to successfully accomplish operations and mission objectives. Task analysis is recognized as an important part of the design phase when changes in hardware, software, and systems are easiest and less costly. Function allocation is also an important part of the design process: deciding whether a particular function will be accomplished by the human or the system, or by some combination of humans and systems. The HSIA risk includes procedure and task design related issues arising from inappropriate definition and development of mission tasks. Operations tempo is driven by the scheduling of mission tasks, and can affect crewmembers' performance, workload, and their SA. The work can be more taxing in combination with other factors such as fatigue, deconditioning, stress, anxiety, and medical conditions. Low workload levels can induce errors associated with boredom that often relate to decreased attention to the task, whereas high workload levels can induce errors that often relate to the narrowing of attention to the detriment of tasks. In addition, when materials such as procedures, directions, checklists, graphic depictions, tables, charts or other published guidance are misleading or unclear, workload is further impacted, and an unsafe situation results. The severity of the consequences may increase with the duration of the mission.

Given that the HSIA is heavily invested in human-system integration, operator tasks, schedules, and procedures must accommodate human capabilities and limitations within the broader system context, and given that crews of LDEMs will experience physical and cognitive changes and increased autonomy, there is a risk that tasks, schedules, and procedures will be developed without considering the human condition, resulting in increased workload, flight and ground crew errors and inefficiencies, failed mission and program objectives, and an increase in crew injuries. The HSIA risk includes this aspect of human-system design as it relates to the definition and development of missions, processes, mission tasks, task flows, schedules, and procedures. To provide adequate task design, we need to understand relevant human capabilities and limitations for performing tasks and how these may degrade on long-duration missions. We also need to understand how other factors may affect human-system performance (e.g., the introduction of automation or advanced procedural aids such as AR headsets).

The procedure and task design aspects of HSIA include 3 contributing factors, which are described below: (a) Requirements, policies, and design processes, (b) Operational tempo; (c) Procedural Guidance.

Task Design: Requirements, Policies, and Processes

A task design that maximizes system performance can be achieved when an HCD process is used during development of the system. HCD is characterized by early and continual user input that is employed in an iterative design-test-redesign process (International Standards Organization [ISO] 13407, 1999), and focuses on making hardware and software, tasks, and related procedures usable by the human throughout a system's entire life cycle. Lack of an HCD process often leads to inadequate task design. Inadequately designed tasks, procedures, and schedules may lead to flight and ground crew errors that threaten crew safety and compromise mission and program objectives. Designs that do not have an adequate level of usability can also lead to lack of efficiency, lack of crew satisfaction, crew frustration, and loss of productivity. Procedures that are illegible, confusing, or at the wrong level of detail negatively impact the task, putting the mission in jeopardy. Schedules that do not take into consideration human biorhythms and physical and cognitive abilities in extreme environments may affect workload that in turn may lead to poor crew performance (e.g., errors or increased task completion time). When tasks are not designed correctly for the human operator additional training is often needed to improve the level of performance. Inadequate task design can therefore result in additional time for training, more support during operations, redesign after deployment, and overall increased life cycle cost.

Understanding the user and the operating environment is important to ensure that design solutions meet the needs of the user within the constraints of the operating environment. This means that full awareness of the user capabilities and limitations, skills and expertise, the work environment's constraints and challenges (e.g., microgravity, isolation, small enclosed volumes), and the tasks and schedule that will be performed to accomplish the mission (e.g., piloting, maintenance, eating, sleeping) is required to ensure correctly specified task designs.

Task analysis is a method within HCD that represents tasks as a hierarchy of steps and actions that are necessary to accomplish goals. It is used to understand and document the sequence of tasks and steps, and the relationship among tasks, to meet the needs of the user performing them. Although recognized as a critical function in design, task analysis is often overlooked until the final design phases when hardware, system, and software designs are too costly to change. It is essential that task analysis be conducted as early in the design phase as possible. Task analysis should be conducted iteratively and should be frequently evaluated throughout the design and development process to allow for proper verification of crew task and system design. Furthermore, task analysis should be performed to identify the "critical" tasks, i.e., those tasks that are necessary to successfully accomplish operations and mission objectives (Kirwan & Ainsworth, 1992).

When iterative task analysis is applied to spacecraft operations it determines operator needs for established mission objectives and concepts of operations. The focus is on humans and the human-system interface and on how humans perform the task within the context of the human-system unit. Task analysis methods for LDEMs may have different issues than methods applied to short-duration missions. Appropriate task analysis methods for LDEMs should be researched and validated. Task analysis of LDEMs will result in a complex list of tasks and subtasks, and a visualization tool should be developed to properly relate these complex systems of tasks, subtasks, and sequences.

Tasks are usually driven by procedures, and when written directions, checklists, graphic depictions, tables, charts, or other published guidance is inadequate, misleading, or inappropriate, an unsafe situation can result. Guidelines for designing task flow, schedules, and procedures are critical to ensure task and mission success. Task support for LDEMs, which can include a combination of visual, video, and verbal procedures and checklists, should be carefully researched and selected so that it aids task performance.

Concepts of Operations (ConOps)

ConOps identifies the characteristics of a system and associated tasks, which in turn integrates the perspective of the human using that system. ConOps and the tasks that comprise them define the operational tempo by which tasks are planned and accomplished. Specific tasks that comprise a ConOps are identified by task analysis. The definition of ConOps requires determining the tasks necessary to accomplish an objective, how they are expected to be performed, and who will perform them (Ainsworth, 2004). When these factors are not considered or addressed appropriately, human performance may be inadequate, and errors may result. Task analyses and ConOps allow for the identification of task features where there is a potential for human error or other issues and problems.

Human Capabilities and Limitations

Considering the human condition as it relates to task performance is critical when addressing and optimizing performance. This includes review of tasks, schedules, training, and procedures to ensure they account for human capabilities and limitations, as well as inevitable physical and cognitive changes that occur during spaceflight. If these and other factors are not considered, impacts to performance can include increased workload, errors, and inefficiencies, and failed mission and program objectives. The severity of the consequences increases with the duration of the mission.

Humans' physical, sensory, perceptual, and cognitive capabilities have constraints that are related to performance inefficiencies, including workload increases and operator error. Regarding cognitive capabilities, for example, the amount of information that can be processed is limited by working memory (Baddeley, 1992; Miller, 1956). Therefore, information overload can be a problem when trying to accomplish tasks that load the working memory of the operator.

On the other hand, information underload can lead to decreased vigilance and can lead to loss of SA, i.e., being less aware of important aspects of the environment needed for the current task and future actions (de Winter et al., 2014). For all these reasons, human capabilities and limitations should be taken into consideration in the design of tasks and associated procedures, hardware, and software. Interactive cognitive aides for use in long-duration spaceflight (e.g. for medical procedures) should be identified, developed, and validated to facilitate crew performance when ground support is limited.

Human capabilities and limitations can be affected greatly by the duration of a mission and the degree of subsequent deconditioning of crewmembers. The strength and aerobic power of crewmembers' load-bearing muscles can decrease during spaceflight missions. On-orbit exercise regiments have been implemented to counteract these deficits, but to date have been only partially effective. Overall, the long-term effects of living in space and the effects on performance are still generally unknown. What is known is that perception in every modality, reaction time, motor skills, and workload can be affected while in space and thus affect performance (Legner, 2004). Thus, it is important to understand how tasks, procedures, and schedules may need to be modified as deconditioning occurs.

Measures of Procedure and Task Design: Workload

Adequate human task performance during space missions depends on usable systems and acceptable workload levels (Gawron, 2000). Workload is the perceived demand on human operators to satisfy the requirements of a given task. Low workload levels have been associated with boredom and decreased attention to task, whereas high workload levels have been associated with increased error rates and the narrowing of attention to the possible detriment of tasks (Sheridan, 2002; Parasuraman & Riley, 1997). If a crewmember is given a system that is difficult to use or they have too many responsibilities to perform, they may become overloaded, and their performance may become degraded. The way tasks are designed and the way they are presented to the crew or are scheduled are critical to the experience of workload. Operations tempo is driven by the scheduling of mission tasks and can affect workload and SA of crewmembers. The same amount of work can be less or more taxing on crew depending on other factors, such as deconditioning or physiological or psychological stress. We do not know how to reliably and unobtrusively measure workload during a LDEM, and there are still many questions to be answered regarding the effects of high levels of intermittent workload or low levels of sustained workload. Workload is an important component of crewmembers' interaction with systems. Designers must consider workload when designing procedures, operations, and hardware and software with crew interfaces.

Because the method to measure workload must match the evaluation objectives, measuring workload for a "system", a set of activities, groups of people, or an entire mission is challenging. The Workload Primer (Casner & Gore, 2010) provides descriptions of approaches to measure workload, and discusses their purposes, strengths, and limitations. This Workload Primer proposes 4 approaches to measure workload for short-duration tasks (on the scale of hours): (1) direct performance measures, (2) indirect measures, (3) subjective measures, and (4) physiological measures.

Operations in space, however, are highly procedural and highly repetitive tasks are completed at the same time every day for extended missions lasting for more than 30 days. This leads to questions about when and what type of repetitive procedures should be automated and the impact this has on operator performance, workload, and team collaboration, and what tasks will become "nuisance operations" because of their repetitive nature and their subsequent interaction with workload. Furthermore, the measure of workload for a system of activities, groups of people, and an entire mission remains difficult to put into operational terms.

Although reliable techniques exist for measuring operator workload (Hill et al., 1992), surprisingly little attention has been directed toward how workload affects performance in extended missions of several months, particularly in extreme environments with expert operators, such as those encountered in space operations. The simplest interpretation of these results might lead one to conclude that workload is best maintained at an intermediate level, in which operators are neither pressed to their limits nor left with their minds wandering.

Given their duration, LDEMs will have a different performance profile than short-duration missions. Astronauts may experience long periods of low workload or outright boredom punctuated with bursts of high workload and emergency and off-nominal conditions. Maintaining a ready engagement among crews during extended periods of inactivity is essential if they are to response effectively to unexpected critical activities. A key factor will be identifying behavioral markers that are consistent with operators who are operating above their maximal or minimal workload threshold during these LDEMs (Canser & Gore, 2010).

Behavioral markers are needed to predict human performance. Cognitive metrics using devices such as electroencephalograms (EEGs) or functional near-infrared spectroscopy (fNIRS) can allow for *in situ* neurophysiological assessments of workload. One fNIRS-based workload tool has been recently validated during completion of flight-based electronic procedures (Bracken et al., 2019), and a wealth of research demonstrates the utility of fNIRS devices to assess workload in real world environments (e.g. Peck et al., 2014). Likewise, EEGs can provide very accurate data relating to task workload (e.g. Hogervorst et al., 2014). Important questions remain, however, on how these metrics could be used to inform changes in tasks and procedure characteristics, and whether they might be effective for adapting real-time user interfaces or procedures.

Automation can reduce operator workload and fatigue. Furthermore, when automation is introduced, the nature of workload shifts from motor (e.g., manual control) to cognitive (e.g., supervisory tasks such as monitoring, projecting, and analyzing) (Parasuraman & Hancock, 2007; Sheridan & Parasuraman, 2000; Billings, 1997). When patterns of workload shift in such ways, new system vulnerabilities arise (NRC, 1993). These vulnerabilities include decreased skill and the human-out-of-the-loop phenomenon (Sarter & Woods, 1997; Sarter & Woods, 1995). As the automation becomes more pervasive, such vulnerabilities may contribute to safety risks. Behavior is unlikely to adapt unless the operators are provided with accurate feedback on their performance. It will be important to know when workload is approaching a given threshold at which performance will be likely to decrease. Feedback would allow the rational human operator to adapt, i.e., to use a shortcut or scrub a task entirely, or to reschedule the task so that it can be completed at a later time. Computational representations of times when workload will be nearing a threshold could be evaluated in simulations before spaceflight, whereas real-time tracking of information regarding workload trends can be provided to the crew *in situ*.

A system that can record and automatically react to real-time changes in user workload would be useful for critical high-workload situations. Additional research would need to be conducted not only on optimizing procedural adjustments to changes in user workload, but also to determine how best to record and use workload data in such a system (e.g. EEG, eye tracking, fNIRS, and other physiological metrics relevant for workload measurement).

Measures of Procedure and Task Design: SA

SA is defined as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (p. 792, Endsley, 1988b). According to Endsley (2000), this understanding and predicting occurs at 3 levels: (1) perception, (2) comprehension, and (3) projection. At Level 1 SA, the human operator must obtain basic perception of critical information. If the human cannot perceptually identify the information needed (i.e., because of poor presentation of information) or does not have access to the information, SA will be severely compromised. At Level 2 SA, the human operator must comprehend the situation by integrating, interpreting, and retaining multiple pieces of information. This means that the operator not only needs to obtain the information needed, but also needs to infer meaning from the information and the relationship of that meaning to task goals. At the highest level of SA, level 3 SA, the operator must predict future events and dynamics in a way that allows for appropriate decisions to be made. According to Endsley (2000), level 3 SA may be reached by highly skilled experts.

SA has been assessed for various types of user interfaces, from pilot displays to power plant control rooms. Methods for evaluating SA commonly fall into one of 4 types: performance-based, knowledge-based, observer-based, and subjective measures.

Knowledge-based measures of SA involve asking participants questions about information related to the current situation and about various interface components. These measures may be provided during a task or at the conclusion of a task. When knowledge based measures are collected during a task, a “freeze” technique may be used in which a simulation is frozen at random times and users are queried about the current situation, or a real-time “probe” technique may be used in which queries are performed concurrently to the task (Salmon et al., 2006). Knowledge-based methods tend to be good for uncovering the declarative (i.e., factual) information that users can identify, but tend to be weaker at uncovering procedural information and decision-making processes (Scholtz et al, 2005). Knowledge-based methods may also be difficult to administer in operational environments without affecting the task performance (Snow & Reising, 2000).

The SA global assessment technique (SAGAT) is the most well-known and frequently used freeze technique (Endsley, 2000). In the SAGAT, a set of questions are derived with the help of SMEs for the 3 levels of SA. These questions are designed to measure the user’s knowledge of the current situation. During an assessment, participants interact with a simulation of the task of interest. At randomly selected times, the simulation freezes and the user’s display displays questions that are randomly selected from the full set of questions. The responses to the questions are analyzed by comparing them to the actual situation at that point in time. The process is repeated several times within a session. The SAGAT has been used to monitor SA during control of simulated autonomous cars (Scholtz, et al., 2005), for pilot threat-detection (Endsley et al., 1908b; Snow & Reising, 2000), and telerobotics for handling nuclear material under different LOA (Kaber et al., 2000). An advantage of the freeze technique is that it can directly and objectively measure the users' knowledge about the current situation. However, it may interfere with the operator's attention (and thus performance) and can only be used in simulations and not operational environments (Salmon et al., 2006).

An alternative to the freeze-based technique is to perform queries during a task without freezing it. As with the SAGAT, SMEs produce a set of questions. These questions are asked at specific periods during the assessment and may be closed-ended (e.g., multiple choice) or open-ended. The answers are analyzed based on their content or the speed at which the participants respond. An example of this technique is the SPAM, which has been used to measure the SA of air traffic controllers (Durso & Dattel, 2004). When true or false inquiries are used, it may be possible to use a signal detection analysis to determine operator SA of a particular type of event (McGuinness, 2004).

Subjective measures of SA require users to produce a self-report of their perceived level of SA. For example, the SART (Taylor, 1990) requires users to rate 10 dimensions (familiarity of the situation, focusing of attention, information quantity, information quality, instability of the situation, concentration of attention, complexity of the situation, variability of the situation, arousal, and spare mental capacity) on a 7-point scale from 1 (low) to 7 (high) after completion of the task. The Cranfield-SA scale (SAS) is another subjective SA measure that was developed specifically for student pilots; items in this scale are based on pilot activities and information that aviation experts determined were critical to maintaining SA during flight (Dennehy, 1997). Although subjective measures of SA, using techniques such as the SART and Cranfield-SAS, have been extensively collected in operational environments, these techniques have been criticized because operators may not know how much they do or do not know, or their scores may be influenced by their performance (Endsley, 1995b). They are, however, relatively easy to administer in operational environments and are non-intrusive (Jones & Endsley, 2000).

Performance-based, knowledge-based, observation-based, and subjective measures of SA each have their advantages and disadvantages, and weak relationships between revealed for these the measures (see Snow & Reising, 2000), which suggests that they may be measuring different SA constructs. Thus, some researchers have suggested using a combination of techniques, such as knowledge-based queries and self-report measures (McGuinness, 2004) when measuring SA.

Measures of SA and crew resource management (CRM; described on page 60) are often intrusive and are inappropriate for spaceflight. However, maintaining effective SA and CRM are critical for avoiding errors, highlighting the need for validated unobtrusive methods and tools to evaluate SA and CRM during future LDEMs.

Measures of Procedure and Task Design: Usability

The International Organization for Standardization has published several standards that contain usability models to evaluate operational usability. ISO 9241-11 defines usability as "the extent to which a product can be used by specified users to achieve specified goals" and recommends evaluating usability in terms of measures of efficiency, effectiveness, and satisfaction. Measures of efficiency relate the level of effectiveness achieved to the expenditure of resources. Measures of effectiveness relate the goals or sub-goals of the user to the accuracy and completeness of achieving these goals. Satisfaction measures the extent to which users are free from discomfort, and their attitudes towards the use of the system. Performance measures can be used to evaluate the effectiveness and efficiency of a system (Tullis & Albert, 2008). The best way to use them is to set up scenarios as described in the methodology section of a usability study.

- Task success

Task success can be easily calculated for any kind of system if the tasks are well-defined. Performance can be scored as "success" or "failure", and confidence intervals for mean success rates can be calculated based on the adjusted Wald method (Sauro & Lewis, 2005). Levels of success can also be defined depending on the nature of the task. Complete success may be achieved with or without assistance, and partial success as well. Failure may be unidentified by the subject, or the subject may give up without completing the task.

- Errors

Errors and usability issues are related in the sense that most errors are caused by usability issues (Tullis & Albert, 2008). Interfaces must be designed to prevent errors from occurring. The lack of prevention points to an existing usability issue of the interface. The researcher must decide on the definitions of errors and severity levels for errors. A very strict error definition could include comments or statements about confusing interface elements (e.g., "I am not sure which button to click"). A more lenient definition may consider only erroneous clicks as errors or the inability to complete a task. Errors, as well as usability problems, have different levels of severity: some errors have more serious consequences, whereas others may be only minor annoyances.

- Task Time

Tracking the time required to complete a task is a good way to measure the efficiency of the system, and is especially informative for repetitive tasks. It can be measured as time elapsed between the start and the end of a well-defined task. Time to complete a task successfully should be analyzed separately from time to complete a task unsuccessfully.

- Efficiency

Efficiency is the amount of cognitive and physical effort required by the task. Efficiency metrics should only consider actions of successfully completed tasks. A metric for efficiency is "lostness" calculated from the number of different pages visited, total number of pages visited, and minimum number of pages needed to be visited to accomplish the task. This metric is most relevant for website usability; however, it may be appropriate for any other interface that requires navigation. Efficiency can also be defined as a ratio of task completion rate and mean task time or number of tasks completed in a unit of time. These metrics are useful to assess a group of users who complete the same tasks, or a single user completing various tasks of comparable difficulty.

- Learnability

Learnability is a measure of system efficiency. One metric, that can be used to assess learnability is change in task completion time over a series of trials. At NASA JSC, the Human System Centered Integration and Design for spaceflight operational processes have used a variety of measures when assessing usability: the Bedford workload scale (Roscoe, 1984); system usability scale (SUS; Brooke, 1996); and an astronaut acceptability rating (not a program requirement, but typically used by the astronaut office). The SUS has been modified to more accurately reflect the astronaut's usability concerns and efforts are being made to increase SUS's "acceptability" and "relevance" for crew use during the Gateway and HLS programs.

Measures of Procedure and Task Design: Crew Resource Management

As distance from Earth increases, it becomes paramount that the crew resources are efficiently allocated across activities. CRM and the more NASA-centric spaceflight-specific term spaceflight resource management (SFRM) refer to how the complex interaction of SA, self-awareness, communication, organizational team dynamics, attention demands, decision making, leadership skills, adaptability, assertiveness, and workload affect avoidance of human errors (Kay et al., 2014). CRM and SFRM concepts and skillsets are used in training to reduce human errors during later operational activities. Metrics and methods are needed to assess the effectiveness of CRM and SFRM when the crew are performing complex activities that require or would benefit from the CRM or SFRM skillsets. Such measures do exist, and as discussed by Kay et al. (2014) may include crew observation, social network analysis, hierarchical task analysis, process mapping, coordination demand analysis, and triangulation. All of these measures can be readily collected in a training environment but are intrusive or time consuming to collect in an operational environment. Therefore, more research is needed to either identify new measures for CRM and SFRM, or modify existing measures so they are less obtrusive and can be administered more efficiently during real operational activities. Such measures would be most useful if they could be interpreted in near-real-time, providing both the crew and mission support personnel with data to guide decision making.

Operational Tempo

Operational tempo is dependent on the definition, plan, schedule, and pace of activities. ISS crewmembers currently rely on ground support teams for most of the planning and scheduling of daily tasks, as documented in the FCI ISS Life Sciences crew comments database. Software tools, such as the onboard short-term plan viewer, provide crewmembers with detailed schedules for daily activities. Although the crew can provide input to these schedules, ground support personnel often adapt and change them. The higher level of autonomy required for future LDEMs will increase the need for automated planning capabilities and tools. Crew input and ground support, as available, will supplement the tasks generated by the automated planning tools. However, automated support for these planning tasks should allow crewmembers to manage daily tasks and ensure that these tasks are performed appropriately when ground support is unavailable.

Developing spaceflight operational plans is a complex endeavor. Planning includes assessing thousands of constraints and uncertain conditions to produce a sequence of commands that will be executed by humans, robots, and groups of humans and robots (Lee, 2002). Plans are optimal when all tasks to be completed are identified and unforeseen events are minimized (McCurdy et al., 2006). Currently, the MCC is heavily involved in scheduling and operations planning. Future planetary surface operations will require that more resources are scheduled by fewer people and that schedules must accommodate increased automation and autonomy. Mission planning will require the flexibility to develop expedited plans for EVAs or for commanding semi-autonomous robots.

Procedural Guidance

Procedural guidance refers to the structure, content, and presentation of procedures. Procedural guidance can contribute to inadequate task design if written direction, checklists, graphic depictions, tables, charts, or other published guidance is inadequate, misleading, or inappropriate, creating an unsafe situation.

Crewmembers are required to obtain, process, and maintain large amounts of complex information to execute spaceflight missions. If this information is not clearly presented the user may misinterpret, overlook, or ignore the original intent of the information, and this can jeopardize performance and risk mission success.

Procedures inform the user on the sequential order of tasks and often provide feedback on the outcome of the task (Salas et al., 2006). Operator error becomes more likely when procedures are written incorrectly or when procedures are overly complicated. Issues associated with procedures that have occurred during ISS missions are directly related to the provision of too much information, lack of diagrams and schematics to illustrate necessary information, and confusion and missed steps caused by multiple links in procedures. These issues have frustrated crewmembers and have directly affected efficient task performance (Rando et al., 2005). Procedures can often complicate or impede the performance of daily tasks, they may call for an inadequate number of crewmembers to perform a task, or specify an inappropriate duration for a task. Procedures are being improved to enhance crewmembers' abilities to acquire information by including more graphic content (e.g., diagrams and images). The goal is to improve the procedures so that they better reflect how operations are actually conducted.

Methods for designing procedures are often based on the collective knowledge and the experience of system designers, simulator trainers, and astronauts to create what is considered the “best” way to mitigate and recover from an abnormal situation or emergency. The current procedures will be difficult to use during LDEMs because of the increased reliance on automation and because of the increased complexity of the systems that are certain to be required for future space missions. Often procedure writers do not understand the integrated nature of the system and its components and do not verify that the procedures for these integrated systems are correct. In addition, there is no assurance that the selected action sequence is indeed the “best” with regard to execution time, mitigation of failure, and maximization of recovery. Because of this, Degani & Heymann (2002) proposed a formal approach for designing emergency procedures and recovery sequences that uses mathematical methods and tools to analyze, verify, and synthesize action sequences, and can be used to verify the correctness, suitability, efficiency, reliability, and safety of an existing or proposed procedure. This approach has been used successfully in the commercial aviation domain, has been incorporated into a new standard for cockpit procedures dealing with in-flight fires, and can be applied to all exploration systems to improve training and procedures (Heymann et al., 2007).

Real-time fault management during ascent and entry ranks among the most safety-critical of spacecraft operations. Consequently, performance of these tasks must be supported by the highest possible quality fault management procedures and interfaces, as well as the most effective choices of allocation of function between humans and automation, within a minimum of display real estate. Two operational concepts for fault management procedures were compared during studies conducted at the NASA ARC by Hayashi and colleagues (2007). The studies resulted in the following guidelines for procedures:

- Wherever possible, minimize the operators’ need to process cluttered, text-rich display formats by automating the elements of the fault-management task that require text processing. If automating those elements is not possible, text-based displays should be structured to reduce searching requirements.
- Conserve display real estate by displaying just the current focus line(s) plus the first 3 lines of the next to-be-completed step in the checklist.
- Allow operators to be able to simultaneously view the electronic procedure viewer and the relevant system summary display format (the display format containing the “soft” system mode reconfiguration interfaces).

Holden et al. (2019a) performed a gap analysis of NASA-STD-3001 by interviewing users of this NASA document, searching the ISS crew comments database, and collecting information during an Autonomous Crew Operations technical interchange meeting. A list of critical keywords related to autonomy was developed. Over 160 relevant documents were identified and obtained. Consensus review sessions were held on human system standards and guideline documents from key agencies, including the Department of Defense (DOD), the Federal Aviation Administration (FAA), the Department of Transportation (DOT), and the Nuclear

Regulatory Commission. Candidate standards and guidelines selected to fill gap areas in the NASA document set have been provided to the Gateway program and to HCAAM researchers.

6 SPACEFLIGHT EVIDENCE

Risk of Inadequate HSIA and Vehicle Habitat Design Factor

Anthropometry and Biomechanical Limitations

Repetitive Stress Injuries
Scheuring et al. (2009) constructed a database of in-flight musculoskeletal injuries that have occurred over the entire U.S. space program. When relevant information was available, they categorized the injuries by type, mechanism, and causality. The data was obtained from multiple sources including postflight medical debriefs, NASA's lifetime surveillance of astronaut health medical record database, and the JSC medical records. A total of 369 in-flight musculoskeletal conditions were identified. Most musculoskeletal injuries involved the hand, back, and shoulders, with abrasions and contusions being the most common type of injury. Although the causality information was limited, injuries were attributed to crew activity, EVA suit, and exercise equipment, and to a lesser extent, the launch and entry suit, the advanced crew escape suit, experiments, EVAs, and egress. There was no specification of whether egress injuries occurred while suited or unsuited. Many of the types of crew activities that Scheuring et al. (2019) identified as causing injuries were related to habitat design: as shown in 5 these included impacting structures (12 injuries), stowing equipment (8 injuries), translating through the spacecraft (8 injuries), repairing equipment (7 injuries), abnormal positioning (6 injuries), transferring equipment (5 injuries), restraint (5 injuries), and donning suit (3 injuries). Improved habitat design could help reduce a number of these types of injuries. The crew activity details were listed as "unknown" for over 16 musculoskeletal injuries, so the number of habitat/vehicle design-attributable injuries could be under-represented. Thus, onboard "memory joggers" are required to inquire about the causes of the injuries, and this information should be made available to designers so they can improve equipment and habitat designs and improve task flows.

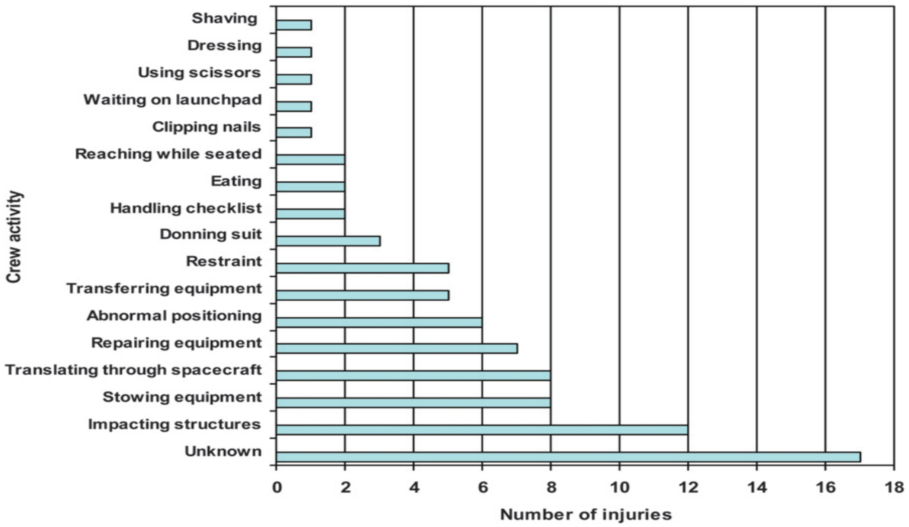


Figure 5. Types of Crew Activity Causing In-Flight Musculoskeletal Injuries Throughout the U.S. Space Program

Although, the FCI ISS crew comments database does not include many reports of crewmembers routinely using awkward postures, there are certain tasks such as stabilizing while doing ultrasound scans, glove box use, and hygiene tasks, that can induce temporary awkward postures. These postures may only occur during the crewmember's initial performance of the tasks, and they learn to accommodate, whereas other tasks, such as some glove box tasks, may involve repetitive motion. Some crewmembers reported that stabilizing their bodies by securing their feet to handrails induced awkward postures, and they sometimes added padding to handrails to improve comfort.

Antonsen et al. (2018) created an accepted medical conditions list (AMCL) that provides a traceable, repeatable, evidence-based consensus process for scoping the medical capability needs for future design reference missions and upcoming programs. This baseline AMCL informs about the high priority medical capabilities that must be included in LDEMs and provides traceable and documented clinical needs to the systems engineering teams tasked with design work. The authors considered 194 medical conditions and their report has a full list of the conditions. Some examples of medical conditions that that are likely to occur, such as nasal congestion, back pain (space adaptation), insomnia (space adaptation), and space motion sickness (space adaptation), have good expected outcomes and require minimal resources. Some examples of conditions where futility and complexity of the treatment are higher are toxic exposure (ammonia), anaphylaxis, neurogenic shock, and cardiogenic shock secondary to myocardial infarction. There is no plan to treat all the conditions listed, but the authors noted those medical conditions that are expected to be treated. This evidence is important to determine the potential injuries that the habitat design could prevent or mitigate.

Impacts of Microgravity on Anthropometry, Posture, and Strength

Changes in body anthropometry have ramifications across all aspects of vehicle and suit design. If parts of the body swell or shrink, the individual is no longer sized appropriately in the suit, increasing the likelihood of fatigue or the chances of injury during suited operations. Thornton and Moore (1987) stated that some crewmembers had trouble donning suits after extended periods in microgravity likely because their torso had increased in length by an inch. Similarly, the fit of individuals relative to the vehicle also changes, causing potential discomfort and restrictions. Thus, current design requirements may not adequately ensure crewmember's accommodation during LDEMs. For example, changes in anthropometry during spaceflight may mean safety hardware no longer fit, or operator errors could occur if biomechanical performance has degraded due to muscle atrophy or other unknown physical changes.

Spinal Elongation

In microgravity, the natural curvature of the spine straightens due to shifts in body fluids and the lack of compressive forces on the spinal vertebrae, elongating the spine and increasing stature. The stature of Skylab astronauts increased by approximately 3% after the first 2 days in microgravity (Brown, 1976). Although this Skylab study involved only 3 subjects, an additional Apollo-Soyuz test flight study found consistent results (Brown 1975; Thornton et al. 1977; Thornton & Moore 1987). Because all of the growth is attributed to changes in spinal length, 3% of the person's stature has been added to the length of the torso (Churchill et al. 1978), and this has driven current requirements in Human Systems Integration Requirements (NASA 2009b). An additional study was completed on board a series of Space Shuttle missions: Stature was measured in addition to seated height to determine if the increase in seated height was similar to the increase in stature. Data was collected from both ISS and Space Shuttle subjects; 29 subjects' seated height and 23 subjects' stature were gathered (Young & Rajulu 2012). Seated height increased by approximately 6%, and stature changes were similar to the 3% increase observed in Skylab and Apollo-Soyuz test flight studies. Results helped identify potential improvements to future vehicle architecture (in particular, seats) and thus potentially increased safety and efficiency of future missions.

Neutral Body Posture

Body characteristics during spaceflight, such as posture, can strongly influence the design of hardware, such as suits. Skylab and Space Shuttle studies have shown that the crewmembers often assume a semi-curled up posture in microgravity, with knees bent and hips flexed (see Figure 6), referred to as neutral body posture (NBP) (Brown, 1975). Posture in microgravity is different from any normal 1-G posture on Earth, and with fatigue and discomfort, the body resists any attempts to force it into 1-G postures. The Anthropometric Source Book Volume I: Anthropometry for Designers (1978) suggests these postural changes should directly drive architectural design for future missions.

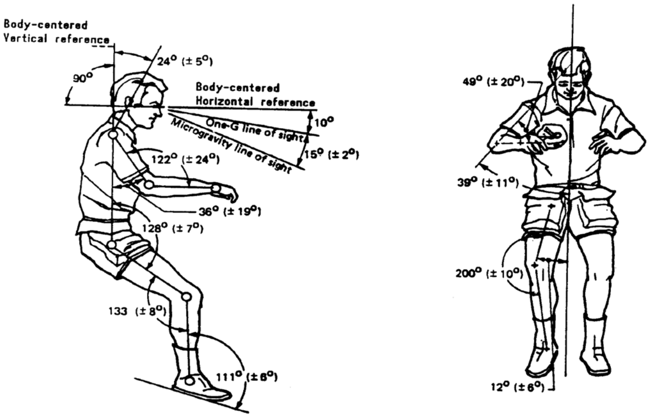


Figure 6. Neutral Body Posture

Mount et al. (2003) found that, in general, 3 primary NBPs were exhibited by the Space Transportation System (STS)-57 crew, all shown in Figure 7. These constituted a slightly pitched forward posture with an extreme bend at the knees (crew 2), an elongated posture with a straight neck (crew 4 and 5), and an almost standing posture (crew 6). The differences in posture could have been a result of the participants' body type, the type of exercise they performed, the amount of exercise regularly performed, or a combination of these factors. Other differences may be attributed to past physical injuries or sex differences such as differences in center of gravity. The duration of the STS-57 mission was 10 days.

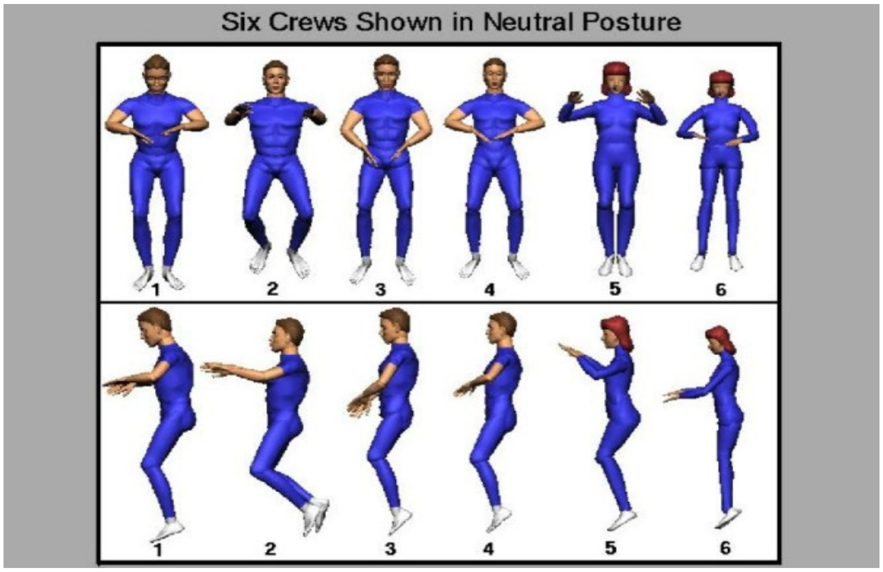


Figure 7. Neutral Body Posture of the STS-57 Crewmembers

Recent studies have collected additional data related to NBP, as well as linear and circumferential measurements of anthropometry during flight (Rajulu et al., 2018). Rajalu et al. analyzed video footage to assess changes in 9 subjects throughout a 6-month mission, and obtained data on circumference, length, height, breadth, and depth of body segments (chest, waist, bicep, thigh, calf) before, during, and after flight. The NBP was also measured to determine changes in body posture (joint angle) throughout the duration of a mission. Overall, the observations in this study are in agreement with the past studies including the Skylab and spinal elongation study. Stature showed a biphasic growth, characterized by up to 3% increase at flight day 15, followed by a steady phase throughout the mission. Other measurements also showed similar trends. For example, increases in shoulder height were moderately greater than increases in stature; however, shoulder height was likely affected by variation in shoulder posture that did not affect measurements of stature. Circumference measurements of the chest, waist, hip, thigh, and calf decreased overall during the mission. The greatest relative change was in the calf circumference, which continually decreased from flight day 15 throughout the rest of the mission to values that reached 10-11% less than preflight measures, and returned close to the preflight measures within 30 days of returning to Earth. The authors stated that this could have be associated with the muscle atrophy, water loss, and other microgravity factors, but interpretation of the results was limited due to issues associated with the crewmembers having to place the markers on the anatomical locations. The authors recommended improving future data collection by enhancing standardization mechanisms and using technologies (e.g., scanner system) to reduce the reliability issues associated with placing markers on the anatomical locations.

Hardware and Task Considerations for Anthropometry and Strength

In the 1990s, some astronauts were disqualified from flying on the Soyuz because they were too tall or too short (Watson, 2007). The Soyuz was ultimately modified to accommodate the U.S. astronaut corps; however, this is an example of the importance of taking population attributes such as anthropometry into consideration during design phases. In some cases, performance of tasks such as disconnecting cables may be hindered during spaceflight due to constraints related to accessibility or obstructions, e.g., there may insufficient room to apply enough strength to complete the task. Some rack areas on board ISS are difficult to access and can impede views to the area when performing tasks such as maintenance. Unavoidable volume constraints on board the ISS can also lead to difficulties accessing and actuating interfaces such as quick disconnects, especially when they require force or torque to operate or manipulate. According to crew feedback in the FCI ISS crew comments database, some cargo straps on a commercial spacecraft capsule required a significant amount of strength to operate because the buckles were oriented at an angle that didn't allow for optimal body position to cinch down the mechanism. If a crewmember must contort their body to manipulate the interface this can increase task duration, impede performance, and induce injury.

Improper actuation force, especially by a deconditioned crew, is also of concern for spaceflight operations. According to the FCI ISS crew comments database, crewmembers have indicated that more force than they expect may be required to mate both contacts on a switch. If a crewmember cannot operate a device, it could create an unsafe environment in the worst-case scenario and, at the very least, will cause increased performance time or error as the crewmember struggles with the device or enlists the help of a second crewmember to help.

The wear and tear on components, as the crewmembers repeatedly use them, is also of concern. The FCI ISS crew comments database includes remarks that locker doors and panel interfaces may come out of alignment, stick, and be difficult to open. Some panels get stuck and require additional tools to open them. If devices are not robust enough to withstand expected use, crewmembers will be unable to operate them or will have to devise alternative means of operating them. The wear and tear on the human-system interface can lead the crewmember to operate it in a less-than-ideal configuration, increasing the potential for injury, equipment or interface damage, and increasing task performance time and error rates while completing tasks.

Consideration must also be made for crewmember size and potential impacts to hardware fit. According to the FCI ISS crew comments database, some crewmembers found the volume of their crew quarters did not accommodate their height or stature and they had to try multiple configurations before finding an angle that would accommodate their preferred sleep posture. A crewmember's size and their preference can determine the need for varying types of restraints for the vast array of tasks performed on board, such as toilet use, science, or medical operations.

Visual Environment

Light Levels for Task Performance

Optimal lighting is important for safe and successful task execution. According to the FCI ISS crew comments database, overall illumination on board the ISS is satisfactory, although some modules provide better lighting and visual environments than others. Considerations must be made for things like stowage that can block and impede lighting within modules. Crewmembers have noted that stowage in the Permanent Multipurpose Module blocks lighting, resulting in a dim workspace. Supplemental portable task lighting is required for some activities such as work completed behind racks and in out-of-the-way locations.

Lighting levels may affect the ability to access equipment and information. For example, according to the Space Shuttle crew reports, during a procedure to reconfigure the downlink rate of the orbiter communications adapter, one crewmember inadvertently selected "BYPBK" instead of "BYPFR" on the rotary knob. This affected photography and TV equipment and other general payload support computer equipment. The crewmember attributed this mistake to the fact that the rotary knob was poorly lit, highlighting the importance of ensuring that displays and controls are properly lit. In future vehicles, where displays may be near sleep areas, proper coverage of bright lighting may be needed.

JSC's Graphics Research and Analysis Facility analyzed the nighttime-lighting visibility of the Shuttle Remote Manipulator System (SRMS), the Orbiter Boom Sensor System (OBSS), and the Space Shuttle tile area during a proposed automated tile scan on flight day 2 of the ISS assembly mission LF-1. To view the joints for the entire SRMS and OBSS and to assess the clearance of the arm to the Orbiter, the available lights and cameras had to be panned and tilted to various areas of joints and wings to view of the critical areas. The preliminary look using the nighttime lighting options available during the automatic tile scan revealed that no single light or combination of lights would allow the entire scene to be illuminated adequately for the cameras. The payload bay lights only aided viewing of the arm directly above the payload bay. When the SRMS and OBSS were primarily at the port or starboard side, only the light emitting diodes on the payload camera could be used, requiring the camera to be pointed where the light was, and generally it could only illuminate one joint or section of the arm or wing at a time. The payload bay doors blocked the view of the area of the wing closest to the payload bay. Further out, areas could only be lit in sections. The reinforced carbon-carbon material along the edge is non-reflective, and when coupled with low light, made clearance viewing extremely difficult with a color television-type camera. Alternative lighting, technologies, cameras, or resources can increase the accessibility of information (Maida et al., 2004).

Currently no data is collected with respect to visibility and lighting conditions during post-mission crew debriefs. No instruments to reliably measure absolute illuminance or luminance have been flown onboard the Space Shuttle or ISS.

Glare and Uniformity

Glare and uniformity are important aspects of lighting design that can affect crew performance. When task areas are not uniformly lit, overly bright or dark areas may become a distraction or reduce visual contrast when reading text. Years of ground-based lighting research has fed into the Illumination Engineering Society Lighting Handbook, which provides detailed recommendations with respect to uniformity, glare, and lighting gradients (DiLaura et al. 2011). Direct and indirect glare can also be an issue, as illustrated by recent HITL studies in Orion vehicle mockups and ISS crew usage of shades.

Color Accuracy

Lighting systems impact perception of color, which may be important for some tasks on ISS. For example, according to the FCI ISS crew comments database, crewmembers reported issues with a task involving comparing the color of a liquid to a color card: the use of a task light during the color card comparison affected the results. This points to the importance of not only providing an adequate lighting system, but understanding the specific lighting conditions for scenarios in which color discrimination is critical.

Blue Light Radiation

Short wavelength visible radiation, such as blue light, can increase health and performance risks to crews. Many recent studies indicate that light exposure of predominantly blue light suppresses melatonin production and can affect sleep (Brainard & Hanifin 2005; Brainard et al. 2001; Brainard et al., 2008; Gooley et al., 2010; Thapan et al., 2010). On the basis of data from HRP-funded research regarding blue light's impact on sleep, new solid-state lighting will be installed on the ISS to replace its fluorescent lighting system. The new solid-state system will incorporate a lighting countermeasure that will control blue light to improve crew sleep.

Flicker

Camera interface problems were encountered with the ISS solid state lighting assembly project: the light source met human factors flicker limitations but ISS camera systems failed to operate properly within the lighting environment provided. Flicker index or flicker limit levels must be established that conform to program specific requirements for human temporal vision, migraine and other light source stimulated reflexes, and must tested with planned camera systems.

System Layout Design

The design and placement of lights and lighting schemes within modules must be carefully considered for the ISS and future vehicles and habitats. Some crewmembers have recommended that lighting always be oriented vertically, whereas others indicate that lighting from the "ceiling" results in dim work areas when they work in other orientations. Crewmembers also have reiterated the importance of providing lighting spares on orbit to accommodate lights as they burn out.

The only lights provided on the ISS are the lights available in modules, and the portable and handheld lights. Lighting fixtures in some ISS modules was not originally installed in a manner that would provide the maximum amount of light output (Baggerman et al., 2004). In addition, lights have failed throughout the life of the ISS and limits on launch mass and volume have prevented the delivery of replacement light fixtures. Lighting in the ISS Node 1 module has been further affected by excessive stowage that has blocked light and reduced the reflectivity of surrounding surfaces, requiring crewmembers to perform certain tasks outside of Node 1, which increases the time necessary to perform tasks and decreases efficiency. The ISS modules are verified and accepted when empty (i.e., no stowage), which may affect the effectiveness of the lighting system.

The FCI ISS crew comments database documents that some crewmembers block light-emitting diode (LED) indicator lights within their crew quarters, using tape or objects on hand such as clothing, before going to sleep. Other crewmembers wear sleep masks. Although not all crewmembers report these actions, the light from LED indicators can disrupt essential physiological processes such as sleep. An example of lighting from displays and indicators impacting the lighting environment is provided in Figure 8. HRP has recently funded a study ("Computational Modeling to Limit the Impact Displays and Indicator Lights Have on Habitable Volume Operational Lighting Constraints"; [PI]: T. Clark) to identify a threshold of light intensity from indicators, displays, and light (both indication and display) that does not exceed established thresholds for a set of operational test cases.

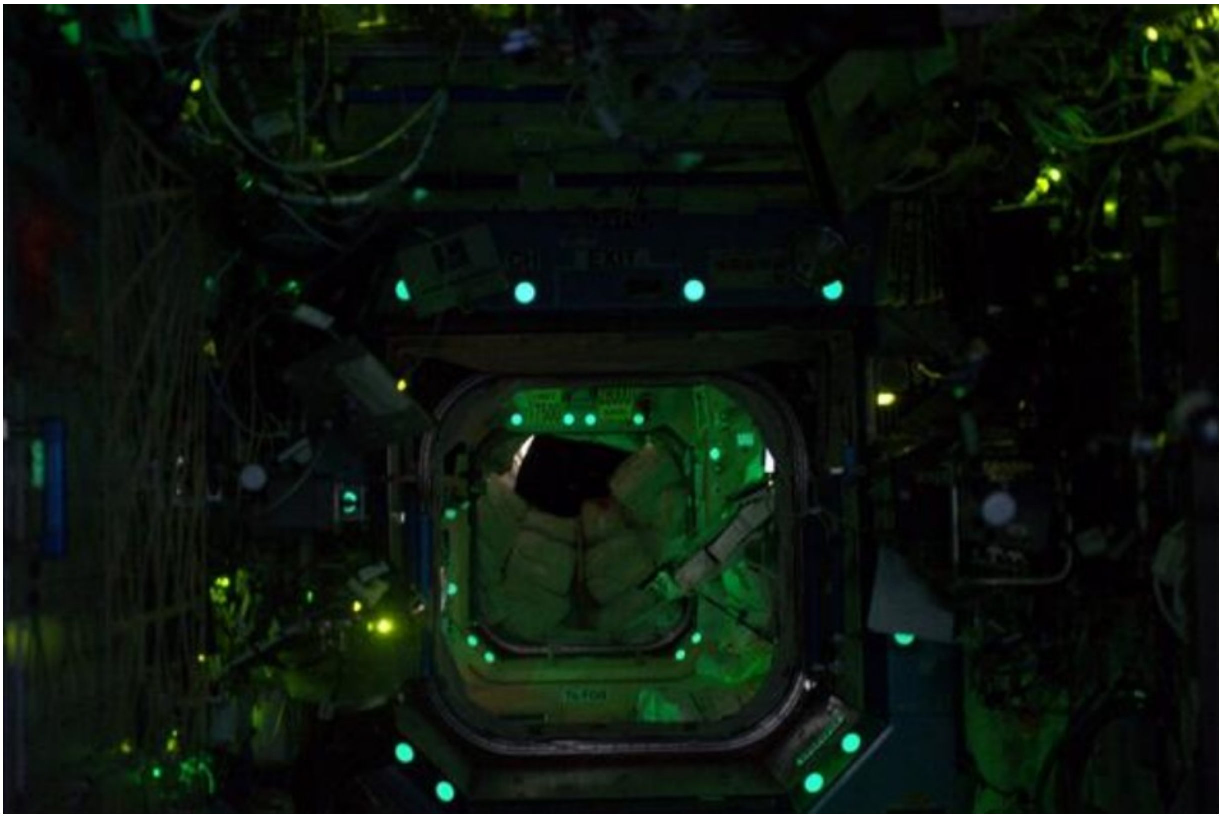


Figure 8. ISS Lighting from Displays and Indicator Lights

The above contributors to poor visibility may result from lack of consideration for lighting system design during the development of concepts of operations or task analyses prior to reviewing the system design. Poor visibility may be due to light source failures and single device failures. Inclusion of visibility and lighting considerations during development of concept of operations and system design, and during task analysis is likely to preclude costly engineering changes later in the program.

Design Consistency

Consistent design of controls and interfaces is important to facilitate efficient operations within the habitat vehicle. Related systems should have a similar interface design, and flight hardware and software should be designed to be as close as possible to the ground systems with which the crew trains (Green et al., 2018). Green et al. (2018) stated that poor consistency across systems, including inconsistency between flight and ground training hardware and software, can lead to errors and increase time to perform tasks. Currently, interface consistency is lacking within ISS, and between ISS and ground based training analogs. Green et al. (2018) presented numerous examples that could confuse the crew, including the lighting interface within the Columbus module, which has 2 different designs of light switches (See Figure 9).

Figure 9. Two Light Switches in the Same Module with Inconsistent Interfaces (Picture from Green et al., 2018)

Green et al. (2018) also provided evidence that inconsistency in the design can affect the efficiency of the crewmember's interactions. For example, having multiple designs of socket and hex heads, and different sizes of socket can increase the interaction time and require more space for different tools. This evidence indicates that updates to standards for consistency of displays and controls are required in the NASA-STD-3001, Volume 2. Revision B.

Readable Labels

Green et al. (2018) reported that unclear or inaccessible labels in the vehicle habitat could lead to errors, safety concerns, and inefficiencies in task completion. The authors noted that when designing labels, system and payload designers should consider the content that is important to the crewmember performing the task, the effects of microgravity (e.g., viewing angles), and how packaging may affect the label readability. The authors stated that a need exists to consider use of colors in labels especially for safety critical tasks in accordance with standards. Green et al. (2018) provides a list of examples of what the crew considers good labels (e.g., labels for the water mist portable fire extinguisher and portable breathing apparatus). Russian labels, unnecessary acronyms, loose labels, and labels that are too small create issues. Figure 10 depicts a hidden label that is hard to read.

Figure 10. Screenshot from Video Showing the Tool Box. The Red Arrow is Pointing at a Label (Picture from Green et al., 2018)

Visual Environment Design Consideration: Controllable Environmental Factors
Green's et al. (2018) report stated that MCC currently controls the temperature of the ISS. The authors note that crewmembers must call down to MCC and request a temperature adjustment, but all crew quarters are adjusted to the same temperature. Given how widely individual preferences vary, this results in situations where some crewmembers are uncomfortable. The crew suggest that in the future, individual crewmembers should have control over environmental factors with each crewmember controlling their own individual crew quarter. Green's et al. suggest that an automated system that customizes environmental conditions automatically based on the preferences of individual crewmembers may improve habitability.

Vibration and G-Forces

Although ground studies of centrifuge-based vibration enable select aspects of visual and manual performance, biodynamic response, and physiological and health consequences to be examined, these studies fail to fully capture the time varying induced G-load and multi-axis vibration environment of actual spaceflight. Centrifuge vibration studies address specific single-axis, single-frequency concerns, and in most cases, both the G-bias and vibration magnitudes were held constant for extended intervals. In addition, all the recent centrifuge vibration studies, and all but one of the Gemini-era studies, were conducted with shirtsleeve participants, i.e., without space suits or helmets. Finally, although various seating configurations have been used, none of the studies have replicated the geometric and structural dynamic attributes of real vehicle seats. These last 2 points are particularly important because they influence where and how much the occupant vibrates. Specifically, the structural dynamic properties of the relatively lightweight seat being designed for Orion will interact with the structural dynamic properties of the seat's occupant, and this interaction will also depend on the conformal geometry at the seat to suit and helmet interfaces.

Although vibration data have been collected for NASA human spaceflight programs, generally, measurements are made for the vehicle itself (see Larsen (2008) for a review) rather than at the seat-occupant interface. Within weeks of the second uncrewed Saturn-V test flight, astronaut tolerance to single-frequency vibration was tested in the lab at JSC (EC/Chief Crew Systems Division 1968) to determine the impact on crew of unanticipated pogo oscillations that were detected at the boilerplate Lunar Module. Although the harshness of the vibration experienced by the test subjects drove a successful re-engineering of the first-stage motors' oxygen line to address the pogo concern before the first crewed Saturn-V flight at the end of that year, Apollo crews still remarked about the severity of vibration from all sources during launch. In contrast, no seat vibration data were available for the Space Shuttle until the end of the program when 3 mission specialist's seats were directly instrumented to record launch vibration as part of an assessment of minimum readable font size on a test placard viewed by mission specialists during 3 flights (Thompson et al., 2010).

Current vibration indicators for Orion and the Commercial Crew Programs are based on coupled loads analysis (CLA) predictions of propulsion system and aero-acoustic interactions of the entire vehicle stack with finite-element models of their respective crew vehicles including seats and displays. Whereas early CLAs suffered from low fidelity representations of occupants' seating (single lumped-mass seating representation), distributed mass models developed by (Carney, 2015) are now employed as part of verification by the various programs to assess and track the health consequences of vibration at the crew-seat interface. In all cases, CLAs are validated (i.e., anchored) by test-flight data from accelerometers mounted on seat backs (or boilerplate seat stubs if the actual seat is not yet available). When outfitted with actual seat-suit-helmet configurations, a centrifuge with time-varying G and vibration-loading capabilities could be driven by the forcing functions determined from test-flight or analytic data to support high-fidelity HITL model validation and operator training, or, if needed, direct verification of vibration related crew performance requirements.

Noise Interference

Data contained in the FCI ISS crew comments database indicate that ISS crews rely heavily on auditory information and warnings. A primary advantage of such auditory information is that crewmembers do not need to be looking at a display to be aware of an alarm and the visual system is free to attend to other tasks. Although noise exposure is an aspect of all living and working environments, the continuous nature of noise exposure from constant sources such as air handling equipment results in a relatively higher noise dosage for crews. Crew health hazards of most concern are temporary or permanent hearing loss, although other effects can be significant. When noise levels are excessive, there is a detrimental effect on face-to-face speech communication, speech intelligibility for radio communications and caution-warning signals, habitability, safety, productivity, and sleep (Grosveld et al., 2003). The ISS acoustic environment is complex and includes many types of noise-generating hardware because the ISS functions as not only the spaceflight crew's home, but also their workshop, office, and laboratory (Rando et al., 2005). The ISS acoustic environment includes continuous noise from the operation of pumps, fans, compressors, avionics, and other noise-producing hardware or systems, and intermittent noise from hardware that operates cyclically, such as exercise equipment or the carbon-dioxide removal system (Baggerman et al. 2004). Some single-event intermittent noises, such as during launch or by fire extinguishing equipment, may be at a high enough levels that hearing thresholds are temporarily shifted for a duration far longer than the intermittent sound itself, although hearing protection is used when possible to avoid these occurrences. Mitigating noise control in spacecraft environments can be challenging and expensive, particularly during later phases of design.

Noise in various ISS modules often exceed the ISS flight rules; levels can be at 67 dBA or higher over the cumulative 24-hour limit (Goodman, 2003). Issues and constraints related to the acoustics environment increase the risk to crew safety because the crew may not be able to hear cautions and warnings (C&W). Although C&W tones are usually audible, there have been a few instances when crewmembers were not able to hear C&W tones due to noise. Noise has also interfered with communication between ISS crewmembers in different modules and between crewmembers in the same module. The FCI ISS crew comments database documents that noise has cost the crew time as they translate between modules to communicate directly, and some crewmembers have reported that excessive noise has negatively contributed to their perception of ISS habitability, e.g, onboard noise has wakened some sleeping crewmembers. Earplugs or noise-canceling headsets can mitigate high levels of continuous and intermittent noise (see Figure 11), but may degrade communications between the crewmembers and between the crew and the ground support personnel. In addition, crewmembers sometimes become uncomfortable while wearing this protection (Rando et al., 2005).

Figure 11. Crewmembers Wearing Hearing Protection Devices and Taking Acoustic Readings

Goodman (2003) states that to achieve optimal acoustic levels during spaceflight should not present a health hazard to the crew; should not present any significant impact or degradation to crew performance and operations; and should provide a habitable, comfortable work and sleep environment. The importance of addressing noise interference will become even greater during LDEMs because crews may be exposed to excessive noise for longer periods of time and will also no longer have the ground team as a backup to monitor for auditory alarms or signals missed on board due to ambient noise levels. Additionally, the impact of self-generated noise from crew within the confined environment can be significant and could be related to reported sensations of "crowdedness" and "lack of personal space and private space" that are cited as key challenges for habitable volumes design (Simon et al., 2011). The negative consequences of lack of privacy are well documented and have implications both for psychological and team health. Lack of privacy has been associated with depressed mood, decreased helpfulness, increased aggression, and a breakdown of community (Altman, 1979; Ashcraft & Scheflen, 1976; Conners et al., 1986; Gallagher, 1993; Newman, 1972; Yancey, 1971). Potential impacts of noise on LDEM crews include increased perceived workload; decreased efficiency of teamwork where voice communications are necessary; increased stress levels; reduced intelligibility; negative impacts to interpersonal voice communication; effects on sleep quality; effects on learning and retention; and subjective impression of degraded habitat quality. Some of these challenges may be met by a combination of noise mitigation techniques for individual components, active noise cancellation, sound masking, variable acoustics, design for acoustic isolation and privacy, and the use of hearing protection devices.

Restraints, and Personal Equipment

Restraints

According to the FCI ISS crew comments database, acclimating to translation along pathways in the ISS is crewmember dependent and can take many weeks to completely achieve. During this acclimation period, crewmembers rely more heavily on handrails placed along translation paths. Although the use of handrails as a translation aid tends to diminish as a crewmember's duration on orbit increases, restraints are still important for tasks that require the crewmember to remain in one stable posture for an extended period of time (DeSantis et al., 2011). Crewmembers have emphasized the importance of restraints being designed to accommodate the operational requirements, the location, and the number of operators for a given task. Restraints that offer greater stability may be required to optimize tasks such as teleoperation and science glovebox operations. For example, to manually teleoperate the humanoid robot astronaut, Robonaut, crewmembers must be stable to avoid any inadvertent movements that would be misinterpreted as a robot command. Restraints may also help accommodate the postural limitations and visual restriction that occurs while performing glovebox and robotics operations in microgravity (DeSantis et al., 2011).

Another important consideration is the ease of setting up and removing restraints to minimize crew time. Some crewmembers feel that they cannot perform robotics operations in the Cupola without foot restraints or foot plates because the Cupola is located over a hatch area with nothing to grip. Some crewmembers require restraints while taking photos in the Cupola because both hands are needed to take a picture. Moving bulkier restraints or portions of a restraint out of a working volume such as the Cupola to accommodate more people within the volume during operations may be cumbersome.

Exposed cooling lines or cables are at risk of damage over time if they are inadvertently used as hand holds. Labels may not prevent crewmembers from grabbing these components when rushing into a space, but they might increase awareness. A 2004 incident on board the ISS illustrates the potential risk: a tiny pressure decay to a braided flex hose that is part of the window system in the U.S. Destiny Laboratory (See Figure 12) was apparently caused by fatigue damage from astronauts inadvertently using the hose as a handhold while looking out the Destiny 20" window (Oberg, 2004; Wilson et al., 2008). The hose extends out like a handhold and is in a location where handholds are needed, resulting in crewmembers using them in an unintended manner.

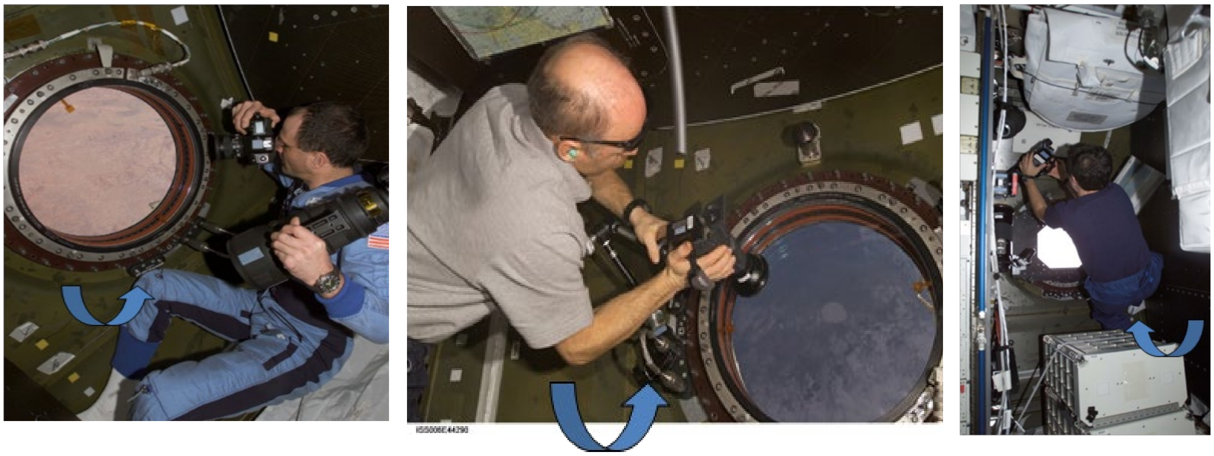


Figure 12. Illustrations of Inadvertent Use of Flex Hose as Restraints and Mobility Aids

Experience with the ISS shows that crewmembers will not use restraints that are overly complex, difficult, or time-consuming to set up or get in and out of. If well-designed restraints are provided, and if the design of workstations, equipment, and task procedures are optimized for the 0g environment, crewmembers' capabilities can approach their capabilities on Earth, and this will also prevent inadvertent use of other equipment as restraints and/or mobility aids.

Crew Equipment

Crewmembers require a lot of equipment to perform their daily tasks, including hardware and tools, computer and network equipment, exercise and medical equipment, and other habitability hardware. HCD is essential to ensure acceptable workload and usability of equipment. If equipment is not designed to be usable by all crewmembers, the likelihood of errors or the inability of the crew to complete a task in a timely manner increases.

As documented in the FCI ISS crew comments database, some ISS hardware items and tools do not have common or consistent interfaces. For example, a mix of metric and English units of measure is used on board because both the U.S. and international partners designed the ISS hardware and tools and no standard was enforced. In addition, some hardware items require unique tools. This can lead to decreased efficiency and negative performance effects.

Accessibility of equipment and tools is also an important consideration within larger habitats. Consistent and efficient labeling of hardware and tools helps to optimize their management, stowage, and use. The inventory, stowage, and management of tools must be continuously addressed to avoid complications related to lost tools or to a lack of necessary and frequently used tools. In addition, different ways to carry individual tools may be required: some crewmembers feel pants pockets don't always accommodate tools properly, and they prefer carrying individual tools in fanny packs.

Although connectors and fasteners are necessary to restrain panels, lockers, and racks, they may be difficult to actuate or may require a particular tool to do so. Some connectors are difficult to mate; their interfaces break over time with use and require additional tools to actuate the broken fasteners. Dzus fasteners are difficult to use, and crewmembers have cautioned they should not be used to secure any panels that contain emergency equipment to avoid delaying or prohibiting access to this equipment.

Visibility/ Window Design & Placement

A key component to optimized visibility in space vehicles and habitats is the provision of well-implemented windows. Windows provide additional onboard lighting, allow the crew to observe Earth and to perform scientific observation and measurements, robotics, rendezvous and docking operations support, and they provide a means of rest, relaxation and improved crew morale. Windows on space vehicles and habitats must allow for successful viewing and imaging. The number and size of the windows, and their placement and accessibility must be considered to optimize design and use. Figure 13 shows a crewmember using a view out of the Cupola before robotic capture of a cargo vehicle.

Figure 13. Astronaut Prepares for Robotic Capture of Cargo Vehicle

To design effective windows, designers must first determine the necessary viewing tasks and the hardware required to perform these tasks. Windows on board space vehicles and within space habitats are often limited, and a multitude of tasks may need to be performed at a limited number of windows.

According to the FCI ISS crew comments database, crewmembers consistently emphasize not only the operational importance but the psychological importance of windows (see Figure 14). The ability to look out of windows is therapeutic, providing peace and serenity, and is important for maintaining morale during long-duration missions. The ISS crewmembers use the Cupola viewing module daily to observe Earth and to take photographs for both pleasure and for research. This window is also used to provide views of docking, EVA, and robotics operations. The Cupola contains a robotics workstation that allows 2 crewmembers to conduct robotics operations while they view operations outside the ISS. Restrictions on the use of existing onboard windows create some frustration among ISS crewmembers, and the provision of as many windows as possible in future vehicles is often emphasized.

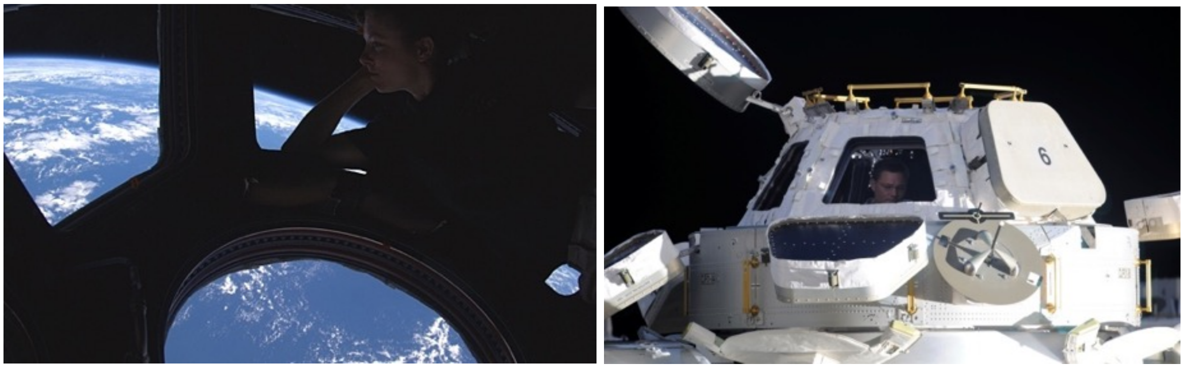


Figure 14. Out of the Window View at Cupola

Green et al. (2018) noted that as much as the crewmembers appreciated the psychological health benefits and the utility of the windows on ISS, they also commented that the scratch panes limited their ability to use the windows for photography. In future vehicles, windows should be coated with material that protects the window but does not affect its utility. Figure 15 is a screenshot from an iSHORT video showing a scratch pane. The NASA-STD 3001 Revision B includes information to support photography in the section on window provision (NASA, 2019b).

Figure 15, A Screenshot from a Video Showing a Scratch Pane on the ISS Window (Picture from Green et al., 2018)

Vehicle/Habitat Volume/ Layout

Currently, the primary source of spaceflight evidence pertaining to habitability and layout is post-mission debriefs conducted by the FCI team. These debriefs, which occur weeks to months after the spaceflight, provide human factors experts with feedback regarding crewmembers' interactions with the vehicle and systems. Some additional information is available from the Journals study (Stuster, 2010)—a study that compiled data from crewmembers' regular journal entries. In the future, additional information will be available from an ongoing ISS Habitability study; for this study crewmembers are using the HRP-developed iPad application iSHORT to capture near real-time feedback about how they interact with their living and working environment (Thaxton et al., 2012).

Task Volumes

According to the FCI ISS crew comments database, most ISS crewmembers found the volume of the crew quarters to be adequate, although the volume of the sleep quarters was deemed too small for changing clothes and to accommodate comfortable sleep positioning for larger crewmembers. However, future exploration vehicles will be smaller than ISS, and it is unknown if the general perceived adequacy of the size of current crew quarters will be relevant for these vehicles. Tasks involving multiple crewmembers require adequate volume for task coordination and team supporting behavior such as body language communication, physical interaction during tasks, and "co-presence" (Nova, 2005). Mutual gaze and co-location also support team cohesion, and enable team members to jointly view relevant information displays, and maintain shared awareness and mental models.

Placement and design of handrails and restraint systems

Figure 16 shows how crewmembers have to improvise foot restraint in the Cupola by using the hatch opening for stabilization. They also explained that in locations where the crew is expected to spend a considerable portion of time, padding on restraints should be provided to increase crew comfort. An example of a padded restraint made by the crew is presented in Figure 16. The crewmembers wrapped a piece of foam around a standard handrail to increase comfort while dining. Green et al. (2018) evidence suggested that there are some updates needed to the Crew Restraint Provision standard and the crew restraint design standard from NASA STD 3001. Vol 2. Rev B. (NASA, 2019b).

Figure 16. Improvised Foot Restraint in the Cupola. The Crewmember is Using the Hatch Opening for Stabilization (Left). Crewmembers Modified this Restraint in the Dining Area to Increase Comfort (Right). (Pictures from Green et al., 2018).

Co-location of Tasks, Zoning, and Topology

Mass and volume of spacecrafts are highly constrained. The living accommodations must be as small as possible while still enabling the crew to accomplish the mission. The co-location and zoning of certain functional habitability areas have been problematic throughout long-duration spaceflight due to vehicle size and topology constraints. The FCI ISS crew comments database provides evidence that placing sleeping quarters adjacent to the waste and hygiene facilities is not optimal because noise made by the equipment can disrupt sleep. Locating dining facilities near exercise equipment and waste collection facilities compromises meal schedules (see Figure 17). Although dining activities can be conducted while other crewmembers are exercising or using the waste collection system, it is not optimal. In addition, locating dining facilities near laboratory work jeopardizes habitability, and compromises the integrity of science activities if foreign debris (such as food products) contaminates the laboratory area.

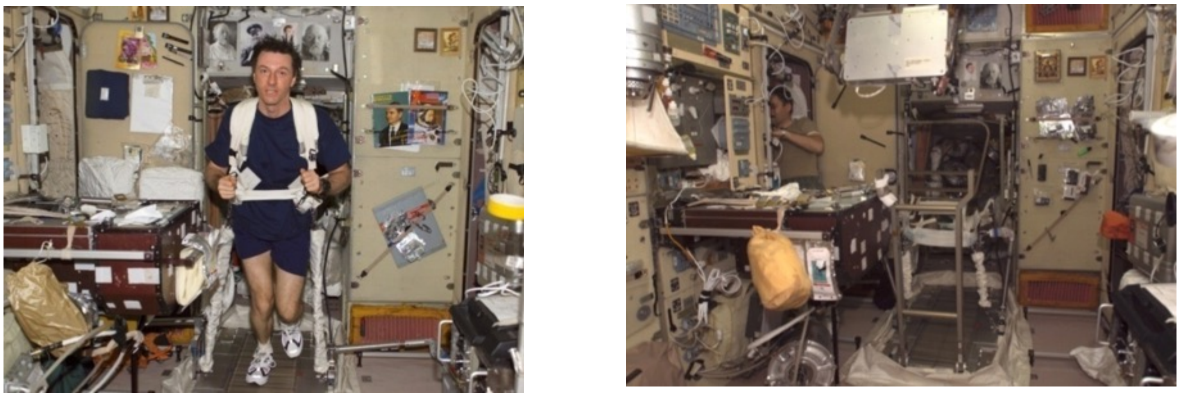


Figure 17. Illustration of the ISS Service Module Configuration with the Galley, the Treadmill, Crew Quarters, and Waste and Hygiene Facilities Co-Located in the Same Habitable Volume.

The FCI ISS Crew Comments Database indicates that several recent crewmembers have preferred to conduct many hygiene operations (e.g., shaving, brushing teeth, sponge baths) outside of the ISS Waste & Hygiene Compartment (WHC), even though the WHC has a designated area intended for hygiene operations This is primarily due to concerns that conducting hygiene inside the WHC could prevent other crewmembers from timely access to the toilet. Some crewmembers choose to change clothes or conduct some hygiene tasks in their sleep stations, but they prefer not to bathe there to avoid releasing water into the space. The fact that crewmembers routinely perform hygiene tasks in alternate areas despite the availability of a designated hygiene cabin area emphasizes the need to consider task constraints due to the co-location of other activities. The ongoing ISS Habitability study seeks to capture more details about this, and as data is analyzed it will be added to future evidence reports. Green et al., 2018 contend that the use of the same volume for both toilet and personal hygiene operations is not feasible because, aside from issues of cleanliness, using the toilet volume for hygiene prevents other crewmembers from using the toilet at the same time. The crewmembers prefer a dedicated private space for personal hygiene, with a full-length mirror to allow for body inspection, and a separate hygiene station provided for each crewmember to stow their personal items. The authors explained that currently on ISS, crewmembers have several improvised spaces for hygiene operations (Figure 18), and do not use the toilet volume for hygiene operations. Some crewmembers commented that changing clothes in the WHC is like changing clothes in a public urinal space, but more inconvenient Crewmembers have to improvise to be able to change their clothes in a different volume. Green et al. provides evidence for the need to update the current standards for Personal Hygiene Capability in NASA (2019b).

The FCI ISS crew comments database includes multiple references to co-location constraints occurring in Node 3 of ISS. Tasks such as personal hygiene and use of toilet, exercising, stowage management, and recreation (Cupola viewing) all in one place can create operations constraints. Volumetric constraints may occur when a crewmember has to translate over another crewmember who is exercising on the treadmill, which is positioned at the entrance to Node 3, and exercise operations on the large resistive exercise device may be constrained if Earth observation or robotics operations are being conducted in the Cupola viewing module. Multiple comments emphasize that the Robonaut in the U.S. Lab can protrude into normal translation paths and require crewmembers to alter translation strategies. Stowage operations may also be impacted when translation paths are blocked by other activities.

Figure 18. Screen shot from Video Showing Translation Path to Improvised Hygiene Area in in the Permanent Multipurpose Module- (Picture from Green et al., 2018)

Accessibility, Stowage, and Clutter

Accessibility problems on the ISS are caused by obstructions and by the design and integration of hardware (Baggerman et al., 2004). The interior components of the U.S. segment of the ISS are grouped into a series of “racks” that were designed to rotate, or tip over, to provide crew access to the rack utility connections and the module wall; however, early ISS crew feedback indicated that rotating the racks is not an effective way to access utilities and connectors in a microgravity environment. The clearance required for human accessibility has been repeatedly cited as an issue in rack rotation capability. The panels and drawers in these racks often “stick” due to the design not operating as intended in 0g, or because too many items are placed in these stowage locations and they are not organized to afford easy operation.

Crewmembers report that the reliance on cabled hardware, and a lack of appropriate cable management, increases difficulty when performing hardware maintenance tasks, prevents hardware relocation, obstructs crew movement and translation, reduces crew efficiency, and in some cases may cause hazards (Green et al., 2018). Figure 19 depicts the excessive number of cables common to many ISS interfaces. A need may be required to adopt more hardware that uses wireless technology. Examples of stowage and cable routing are shown in Figure 19. Operations are impeded if stowed items cannot be easily located or identified.

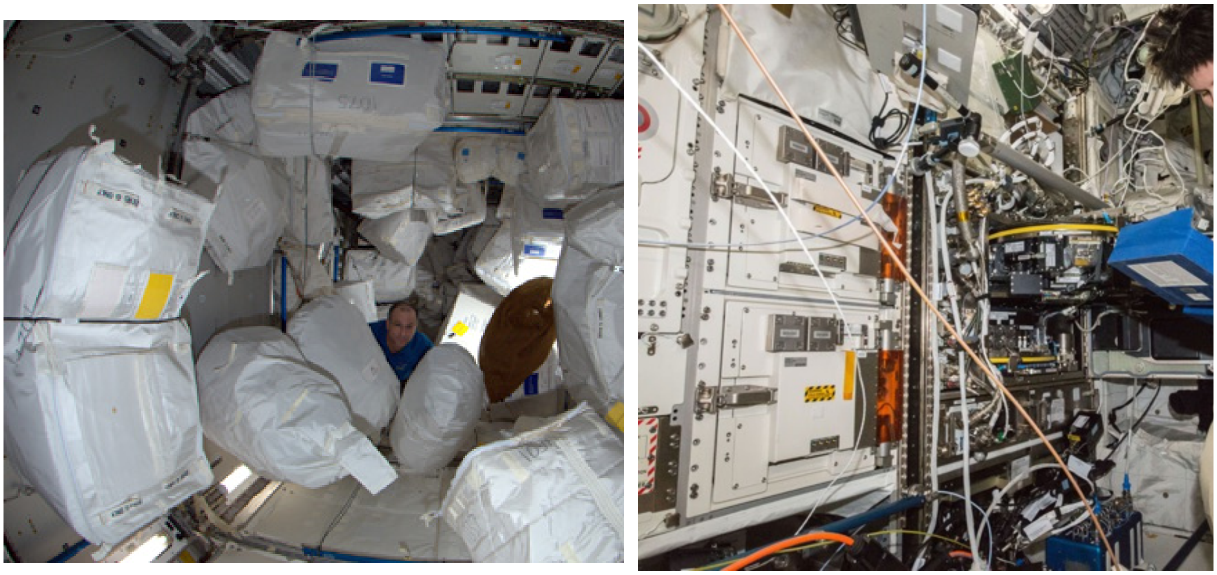


Figure 19. Left: Stowage in the Permanent Multipurpose Module. Right: Example of Cable Routing on ISS

As stowage accumulates on the ISS, some items are stowed in front of panels and in translation paths, reducing the crewmembers’ ability to access items quickly. Increased onboard stowage and the need for additional methods to manage stowage has led to the recent introduction and use of radio frequency identification tags onboard the ISS to track stored objects. The effectiveness of this technology is currently being evaluated by the ISS crew.

Green et al. (2018) reported that contaminants such as fungi and bacteria could lead to unhealthy conditions, and that dirty surfaces can affect the comfort and well-being of the crew. The authors suggested that interior surfaces should be constructed of easily cleanable material, especially areas that are likely to get dirty (e.g., in the galley and waste and hygiene compartment). Figure 20 shows a dirty surface in the dining area that concerned the crew.

Crewmembers must be able to easily access air filters and other areas that collect dust without the use of tools so they can perform regular cleaning that should improve airflow. Figure 21 shows examples that the crew defined as good and bad examples of grill design.

Figure 20. Surface Dirt Accumulation in Dining Area (Picture from Green et al., 2018)

Figure 21. Good Grill Design (Left); these Grills are Highly Visible and Easy to Access and Clean. Poor Grill Design (Right); This Grill is Hard to Access for Cleaning and Requires Tools.

Stowage accumulation on the ISS has been exacerbated by the buildup of packing materials that arrive with each shipment (currently by international partner resupply vehicles, such as the Russian Progress Module or commercial resupply vehicles). Limits associated with the ability to dispose of packing materials results in excessive amounts of stowage space being used for waste. All designated stowage areas on the ISS are now full, and items are being stowed in areas intended for habitability and work-related functions, in passageways, and in front of other stowage areas. In some instances, the stowage violates the allowable limits for the habitable volume areas. When crewmembers are searching for items, they must move many other stowed items out of the way to gain access to the desired item. During some ISS expeditions, items have blocked emergency fire ports, illustrating the risk that excessive stowage can pose on the crew's safety and efficiency. Green et al. (2018) reported that stowage issues affect crew mobility and operation of payload containers. Figure 22 shows the inconvenient of de-mating a payload container lip, moving the surrounding connections, and trying to open the lip 90 degrees to slide a payload out.

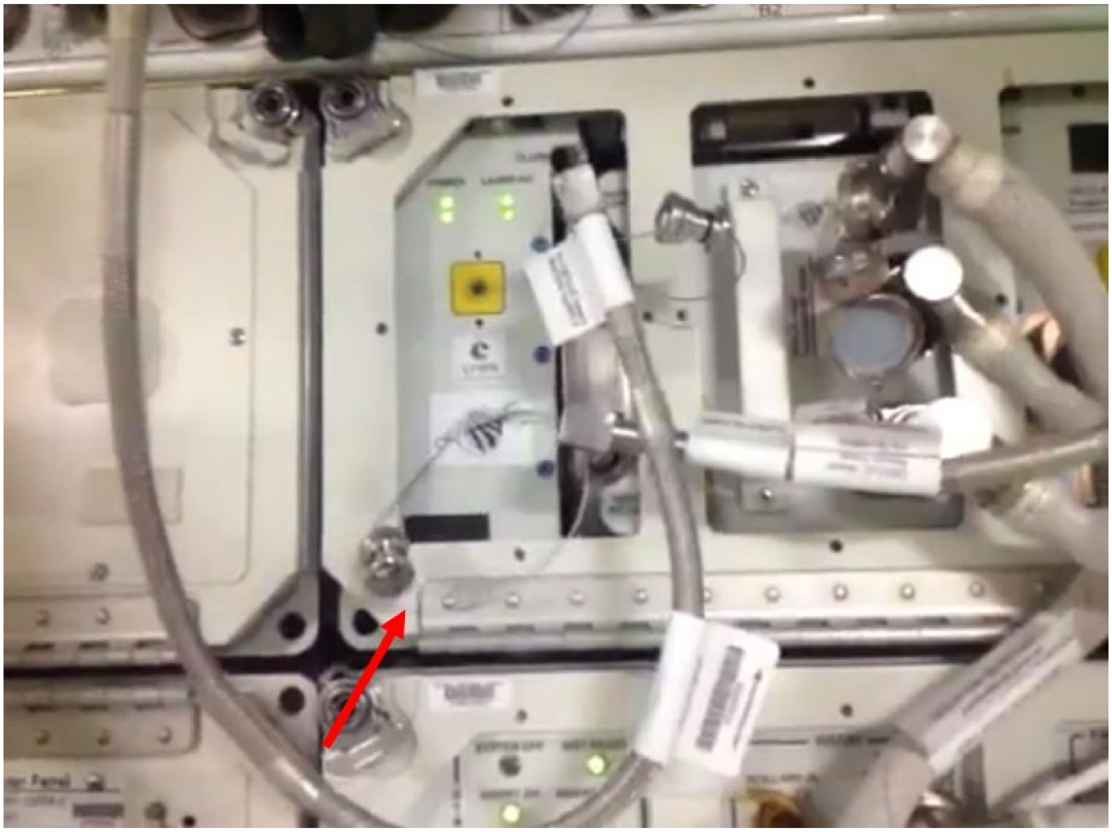


Figure 22. Screenshot from a Video Showing the "Lip" Over Payload Containers. To Access the Container a User Must De-mate the Connections and Open the "Lip" (Picture from Green et al., 2018)

Green et al. (2018) recommended stowage systems be designed to accommodate the specific tasks constraints and the location in the habitat where the tasks take place. For tasks performed frequently and/or for emergency tasks, time to stow and unstow materials may delay and frustrate the crew. The authors recommended that certain pieces of hardware should be considered for temporary or permanent stowage near the activity site. Different forms of stowage (e.g., clear and opaque bags of various sizes, compartments, and clasps) should be developed, and some control should be given to the crew. For example, crewmembers may have different stowage requirements and preferences for their private crew quarters, which should be accommodated when possible. Stowage systems (lockers, cargo transfer bags, etc.) should be easily accessible with microgravity operations in mind. The evidence provided by Green et al. (2018) suggests that the standards for stowage provisions and stowage interference in the NASA-STD-3001, Vol 2 Revision B (NASA, 2019b) should be updated.

Another important aspect of stowage management is balancing the launching of ISS supplies (manifest) with the ability to dispose of waste and return items to Earth (down mass)

(Baggerman et al., 2004). The habitability of the ISS was affected by an imbalance between the mass limits of launch and return when the Space Shuttle was grounded after the Columbia accident because no systematic approach existed to dispose of unused hardware and supplies. Because onboard stowage has, at times, exceeded the allowable ISS requirements in the habitable volume, stowage levels are constantly tracked and evaluated. The onboard inventory of supplies (such as clothing and hygiene supplies) has increased over time, and these supplies continue to be manifested: each crewmember brings some personal items with them to the ISS and at the end of their stay the unused items remain. This increase in inventory contributes to crew safety risks because items are placed in the habitable volume and translation paths. This situation has improved somewhat because of better inventory management and streamlining of the manifesting process. However, the inconsistent nature of disposal capability and inconsistent practices for tracking hardware and supplies remain problematic.

Although ISS crewmembers have provided generally positive comments regarding trash and odor control, trash is stowed on the ISS until disposed of on a departing vehicle (Green et al. 2018). Depending on the frequency of visiting vehicles, the trash may be stowed for long periods of time and crewmembers have reported smelling bad odors after a few months (Green et al., 2018). These bad odors increase if there are partially closed bags. Future LDEMs that have no visiting vehicles will require a trash system that minimizes impacts to the habitable volume. Green et al. (2018) contend that this evidence illustrates the need to update the NASA-STD 3001 Revision B for LDEM.

Improvement of package designs can mitigate trash stowage issues and make processes more efficient. Green et al. (2018) reported that complex or inefficient packaging or stowage of material adds time and frustration to task performance, and excess packaging becomes trash that occupies valuable volume. For example, crewmembers reported that some equipment and food have multiple packages that contribute to the accumulation of trash, and operation time increases when they have to open multiple packages. Green suggested that a task analysis could determine the most efficient way to package material based on the expected tasks, and they suggested research on alternatives to standard packaging materials (e.g., reusability or degradable materials) and requirements to maximize crew efficiency.

Habitat Configuration for EVA Operations

Habitat configuration should not only be addressed for activities inside the vehicle, but also for internal EVA spaceflight activities, i.e., when a module of the spacecraft is depressurized and suited crewmembers access and repair the module. Habitat configuration is a key to maximizing task efficiency during EVAs because EVA introduce additional risk. Off-nominal maintenance on board the Mir space station provided some insight into safety risks related to on-orbit operations and habitat configuration. When a Russian Progress module collided with Mir, the Spektr solar array and thermal control system radiator were damaged, which led to the depressurization of the Spektr module. Crewmembers had to enter the depressurized module wearing space suits to perform repairs, reconnect the electrical power lines, and conduct detailed safety and operations assessments. There was limited area within the habitable volume for the suited EVA crewmembers to move and work inside the module and hazards included sharp edges, fragile materials, fluid contamination, touch temperature issues, and entanglement hazards from fans. An internal EVA hazard assessment was conducted to identify and mitigate the hazards, and procedures and training were developed to ensure the crew could safely conduct the operations, leading to successful completion of the internal EVA. Although the configuration of this work environment was unique, it is still important to consider potential issues with habitat configuration in these types of off-nominal situations to ensure crew safety and efficient task performance.

Habitat Integration

When a task or operation requires multiple components that are not integrated, crewmembers become frustrated and dissatisfied (Green et al., 2018). Further research is required to integrate systems within habitats and vehicles to increase crew efficiency and feelings of work satisfaction, and to decrease feelings of frustration.

Behavioral Impacts on NHV

Stuster (2010) analyzed the journals that 10 crewmembers composed during their ISS missions (up to 215 days), and assumed that participants would write more frequently about topics that are most salient to them. Topics addressed the journal entries were sorted into 24 categories: “equipment” was the sixth most frequently discussed category (mentioned by each crewmember an average of 50 times per mission), and “habitability” was most frequently discussed subcategory within “equipment” (mentioned by each crewmember an average of 3 times per mission). Although some categories were mentioned less frequently as the mission progressed, “habitability” was mentioned most in the first quarter of the mission, less frequently during the middle of the mission, and mentioned more in last quarter of the mission to levels approaching that seen in the first quarter. A similar trend is seen with the subcategory of “stow/restow”, the most frequently mentioned topic in the "logistics/storage" category. Essentially, the prominence of habitability does not seem to diminish over time, and in fact if left unresolved, may present a stressor that accumulates with time.

Crewmembers’ comments gathered during the Journals study also demonstrate the relationship between environmental design and behavioral factors, as shown in the examples below, the last of which affects team cohesion:

“I will mention a relatively minor thing that is starting to become an irritant. That is the high-pitched loud noise that comes from the rack. I’m pretty sure it is the pump and it is way too loud. One of the environmental things I miss the most is quiet—complete silence. Unfortunately, my primary workplace is right in front of it because of its proximity to the panel and the laptop I use most.”

“So many laptops have failed that we now have to scavenge. Psychologically, this is significant because I am now effectively computerless in the Service Module, so spend my browsing time in the Lab away from (other crewmember). It feels strange—only meals and the toilet lead me to the Service Module now."

"There's only room for 2 at the table, so 3 of us just float and juggle our food. I'm not sure what the idea is for when we have 6 of us up here."

The insight into habitability gained from this study aligns with the types of issues identified during crewmember debriefs, and highlights the importance of habitability, not only in terms of crew efficiency but also its effects on behavioral health of individuals and teams.

Layout Re-configurability

Habitable volume can be maximized if spaces can be re-configured. Crewmembers have reported that the ability to manually re-configure the layout of a space may help them perform certain tasks (Green et al., 2018). Software can also be used to re-configure system operations, but crewmembers will not use this type of software function if doing so requires more effort than a manual re-configuration of equipment. Re-configurability must be accomplished with seamless interfaces that allow timely and effective results.

Risk of Inadequate HSIA and Human Computer Interaction (HCI) Factors

Informational Resources/Support

In 1969, on Apollo 10, confusion caused by poor mode information display and crew communication issues resulted in the capsule spinning out of control. The crew inadvertently switched the command guidance system twice to the incorrect mode. Due to lack of any proper feedback from the system, the crew was unaware of their mistake and tried other troubleshooting measures that further escalated the problem. Had this not been corrected, the capsule would have impacted the moon, killing all on board (Shayler, 2000).

The ISS is one of the most complex human-machine systems ever created. Informational and resource support challenges are a continuous concern. Several communication issues, documented in the FCI ISS crew comments database, include miscommunications, unrealistic demands, ineffective interpersonal communication techniques, and a lack of understanding of on-orbit life. Deficient communication between the ground and crewmembers can cause frustration and negatively affect performance. This can be due to ground operators miscalculating the time required to complete a task. Many times, crewmembers could not interpret the information provided by the ground support personnel, or determine what tasks could be automated to help with productivity (FCI ISS crew comments database).

ISS crewmembers rely on procedures written by ground personnel to support and guide onboard operations. Crewmembers have reported issues such as some procedures being too complex, lengthy, and difficult to follow due to too much information. Information has not always been presented in a usable format and some procedures lack diagrams and schematics to illustrate necessary information. These issues have frustrated crewmembers and directly affected their performance. Procedures are currently being improved to enhance crewmembers' abilities to acquire information and better support operations by including more graphic content and improving information presentation to avoid unnecessary information or missed steps. For example, crewmembers suggest that when steps must be skipped, instead of including instructions about skipping the steps these steps should be crossed out on the actual procedure, eliminating the need to go back and forth between the instructions (about how to follow the procedures) and procedures (tasks to be done by the crew). These changes, which may decrease errors and increase efficiency, are and being implemented in more recent updates to procedures (FCI ISS crew comments database).

Many commercial applications such as Microsoft Word and Outlook are installed on the Payload and General Support Computers (PGSC). Other PGSC applications are less commonly used or are proprietary and relatively inaccessible to crewmembers other than during simulations and on training laptops, which increases preflight training requirements or can affect on-orbit operations. For example, during one mission, crewmembers had difficulty using a graphics viewer/editor, and the operating system would unexpectedly reconfigure PGSC settings. This caused significant problems with some software applications. Shuttle crewmembers have reported that they are not provided with enough training to diagnose or correct this problem. Thus, poor informational resources and support for using and troubleshooting PGSC software, has led to crewmembers' frustration.

Attention and Cognition

The Space Shuttle Crew Reports document how the design and placement of computer systems can affect appropriate allocation of attention. In one instance, crewmembers using the Dynamic Ubiquitous OnBoard Graphics (DOUG) Program for dual-arm operations had a near-miss between the Space Station Remote Manipulator System (SSRMS) and an ultra-high frequency antenna. Although the joint angles of the Shuttle Remote Manipulator System (SRMS) were fed real-time into the program, the SSRMS positions had to be input manually, a task that required attentional resources. If real-time SSRMS joint angle information had been supplied to the program, then the near-miss between the SSRMS and antenna might have been avoided. Additionally, crewmembers suggested that a warning on DOUG when a robotic arm gets close to structures would have helped them focus their attention on tasks to prevent a collision.

During spaceflight, crewmembers may experience a condition called asthenia, a general sense of weakness. Anecdotal evidence exists that asthenia, among other things, leads to decreased memory, attention issues, and a general increase in sensitivity to external stimuli that may result in false alarm behavior (Aleksandrovskiy, 1976). NASA conducted a review of the literature on asthenia and concluded that further research is needed (Sandoval et al., 2012). Given some of the symptoms of asthenia, it should be an area of concern for HCI development.

Some crewmembers have deficits in grammatical reasoning before launch and after landing (Manzey et al., 1998), which suggests a general slowing of their cognitive processing speeds.

Furthermore, elevated CO_2 levels are associated with reduced grammatical reasoning (Manzey & Lorenz, 1998). How these changes in grammatical reasoning affect an individual's interaction with the grammatical components of displays and instructions and their resulting performance is unknown. Displays and instructions may need to be tailored to compensate for decreased grammatical reasoning during spaceflight.

During spaceflight, crewmembers can have cognitive issues related to sleep deprivation. Basner et al. (2015) have developed and validated a cognition test battery for spaceflight that provides 15 unique forms of 10 neuropsychological tests covering a range of cognitive domains that include sensory-motor speed, abstraction, working memory, spatial learning and memory, spatial orientation and concept formation, emotion identification, abstract reasoning, complex scanning and visual tracking, risk decision making, and vigilant attention. This cognition test battery (Basner et al.,2015) could be used to further study the effects of LDEMs.

Memory impairments are a concern during spaceflight (Strangman et al., 2014). A study conducted before, during, and after a 438-day spaceflight found that performance on the Sternberg (1966) memory-search task was reduced in the days before launch and after landing (Manzey et al., 1998). Search speeds declined most noticeably 2 and 3 days before launch, recovered during spaceflight, and slowed again during re-adaptation after landing and remained slowed for at least 12 days. The memory task used in the study specifically tested short-term memory of sets of letters. How this generalizes to short-term memory performance during mission-related tasks is unknown. Interface-based memory aids may be required to ensure appropriate performance after landing on any terrestrial body after spaceflight. Dual-task performance also decreased in the days before launch and after landing (Manzey et al., 1998). Given the complexities of spaceflight, multitasking may be unavoidable, so displays and interfaces should be developed to help users manage multiple simultaneous tasks.

Motor skills/Coordination and Timing

Correct sensorimotor coordination relies on several functions to integrate information from the environment relative to body position and orientation. These functions must adapt to gravity alterations (introduction to weightless, reintroduction to Earth gravity, or introduction to surface gravity levels of other planets). These periods of adaptation can give great insight into the mechanisms of sensorimotor functioning.

Anecdotal reports (Barratt, 2012) suggest that astronauts undergo an "adaptation phase" during the early part of space missions, in which new movement strategies are developed in response to the microgravity environment (Manzey et al., 1998). Similar adaptation phases have been found for space motion sickness (Barratt, 2012), cardiovascular regulation (Ortega & Harm, 2008), and psychological functions and sleep (Hamilton, 2008).

The idea that microgravity would influence perceptual-motor performance seems valid, given the influence gravity has on the vestibular and muscular systems (Lackner & DiZio, 2000). A number of sensorimotor functions depend on gravity, such as postural balance, hand-eye coordination, and spatial orientation (Clément, 2007). The vestibular system responds to linear and angular accelerations of the body; these responses then become integrated with visual and some esthetic inputs to generate the appropriate muscle movements. During periods of adaptation, sensory inputs that could be misinterpreted, leading to incorrect body responses and could threaten crew safety or mission objectives. For example, one cosmonaut during the early phases of an 8-day mission on Mir had performance impairments using a joystick during a tracking task. A second cosmonaut during a 438-day mission on Mir had similar impairments, but because of confounding variables, microgravity could not be established as the sole contributor to decreased performance (Manzey et al., 1998; Manzey et al., 1993, 1995). However, during a 20-day mission on Mir another cosmonaut performed the same tracking task with the addition of an aiming component, and impairments were attributed to an underestimation of the mass of the arm and hand caused by exposure to microgravity (Manzey et al., 2000).

Four astronauts on an 8-day space shuttle mission (STS-89) took longer to move a joystick or a trackball during an aiming task than they did on the ground (Newman & Lathan, 1999). In contrast, 2 studies measuring visual-motor coordination of 6 astronauts during a 16-day Neurolab shuttle mission (STS-90) revealed no performance decrements attributable to microgravity (Bock et al., 2001), possibly because the astronauts were stabilized with a harness during the task. A study of 7 cosmonauts on Mir found that an inability to see their hands during a tracking task did not influence their performance (Mechtcheriakov et al., 2002), suggested that slowing of the movements was an adaptive response to preserve accuracy in weightlessness or because of body's instability in the lateral plane. Another study of 5 astronaut's (STS-117 and 118) performance of a tracking and aiming task with a single- and dual-task component (Bock et al., 2001) found decrements in performance only with the dual-task paradigm, suggesting cognitive overload, rather than microgravity, was the primary factor in decreased sensory-motor performance.

Coordinated or timed movements require information from the visuo-motor system. People sense patterns in gravitationally governed events by their ability to recognize the dynamic properties like special scale (or size) of that event (McConnell et al., 1998). This suggests that by perceiving the kinematics of an object (for example trajectory), people have intrinsic knowledge of how the underlying dynamics of the object is governed by physical laws of gravity (Runeson & Frykholm, 1983). When these laws become violated, such as in microgravity, this could lead to inaccurate perception of the dynamics of the object (e.g., moving or falling objects, or collisions between objects) and could change coordinated or timed actions of the user.

This hypothesis was tested on a Neurolab experiment in which astronauts were asked to catch a ball projected with various speeds (McIntyre et al., 2003). In normal gravity (1-G), a peak of anticipatory biceps activity occurs in about 40 ms, and the forearm rotates upward to meet the ball and the hand stiffens. These events can tell us when the brain expects the ball to arrive.

Results showed that the timing of an Electromyogram peak and forearm activity started too early in zero-G, with the limb movements stopping and, in some cases, reversing once the error in anticipation was perceived. This suggests that neural responses can be corrected by updating estimates of time-to-contact based on visual feedback. During microgravity and gravitational transitions, displays that provide such visual feedback could improve neural responses, and improve or correct perceived projection and speed of objects.

Fine motor skills are important when interacting with displays, controls, touchscreens, and gesture-input devices during spaceflight. During a research investigation on the ISS, 7 astronauts completed a battery of tests before flight, over the course of a 6-month mission, and out to 30 days after landing to determine how long-duration exposure to microgravity and gravitational transitions affect fine motor performance (Holden et al., 2018a). For this study, a touchscreen tablet-based fine motor skills test battery was developed (and tested with the gold standard Pegboard task [Thompson et al., 2015]). The test consisted of pointing, dragging, shape tracing, and pinch-rotate tasks (see Figure 23).

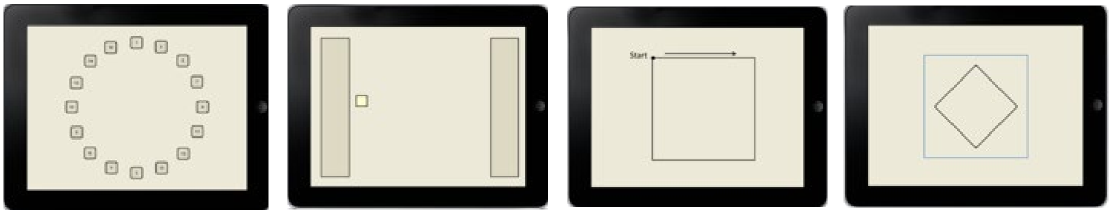


Figure 23. Fine Motor Skills Battery Tasks

Holden et al. (2018a) found no significant decrements in performance during the mission, and small but significant decrements at the gravitational transition points. However, the performance on the pointing task after flight was worse than the performance of ground controls, even 30 days after landing, indicating that gravitational transitions are detrimental to fine motor performance, and the effects may last longer than initially anticipated. Similar effects were observed by Moore et al. (2019) in their study: astronauts experienced a decrement in motor function and motion perception after 6-months aboard the ISS as determined by a significant decline in manual dexterity and a degraded manual tracking performance during dual tasks. Furthermore, these decrements in performance were not due to fatigue as was confirmed from the results of a sleep deprived control group on the ground. This study also found that astronauts had a significant impairment in their driving ability (on a simulated driving task), which raises concerns about future astronauts landing on a planetary surface and performing tasks such as driving a rover. Astronauts may also need to perform safety critical operations after landing that would require fine motor skills. Thus, there is a need for more research in this area, with studies of different durations incorporating different post-landing tasks and crew-like subjects. Developing CMs will not be possible until fine motor performance is better understood.

Environmentally Induced Perceptual Changes

Although reports are anecdotal, the Space Shuttle Crew Reports have documented that during proximity operations, pilots have reported that ISS has appeared closer and the closing rate faster than was indicated by state vectors, an effect that may be due to environmentally induced perceptual changes.

Regarding the ability to see the Space Shuttle displays during launch vibration, some crewmembers claim to have no difficulties, and others say there were points in time during which the displays were unreadable. Effects of vibration on human performance became a high priority when Orion spacecraft development began.

Tri-axial vibration of seats was measured during launch of the Space Shuttle flights STS-119 and STS-128 (Adelstein et al., 2009a) and the vibration effects on visual performance was assessed during launch by having the mid-deck crewmembers examine a spacecraft system display printed on a placard. The crew examined 4 font sizes to evaluate readability at 5 phases of vibration (Thompson et al., 2010). Not surprisingly, the results show that, in general, as vibration increased font sizes required for readability increased. Font size limits for Orion were determined based on these results. Seat headrest vibration in x- and y-axis appears to have the greatest impact on visual performance, and seat vibration in the z-axis has less of an impact (Thompson et al., 2010). As propulsion technologies change with the advent of new space vehicles, we may need to investigate the effects of new vibration profiles on crewmembers' ability to read and operate human-computer interfaces.

Design of Displays and Controls

During Apollo 11, the Lunar Module used a computer system that displayed output only in codes and lacked an intuitive interface. Many codes were memorized during training and simulation on the ground, but during the first Apollo landing, a code appeared that had not previously been encountered in simulation. The crew had to decide whether to proceed with landing or abort, and so ground personnel were consulted. After a short pause, the ground responded that the code was a non-critical alarm code and that the crewmembers should proceed with landing. Thankfully, the landing was a success (Jones, 2012). However, this incident highlights a critical concern as we move toward more autonomous missions where there will be communication delays or in some cases blackouts. Computer interfaces for LDEMs must provide enough detail, context, and priority to allow crewmembers to autonomously make safety- and mission-critical judgment calls.

The FCI ISS crew comments database documents several issues associated with the usability of displays. One frequent issue is the lack of a common overall infrastructure and layout to promote ease of use and understanding of intended operations. Valuable ISS crew time has been lost as crewmembers struggle to understand the use of disparate ISS displays. Inconsistencies in ISS displays have led to incorrect data entry, navigational errors, and inaccurate interpretation of the data in the displays. When display interfaces are dissimilar and information is not presented consistently, crewmembers may require additional training and time to master the displays. The crewmembers also incorrectly generalize from one display design to another. These errors can compromise crewmembers' safety, especially in the event of an emergency.

The Human Research Facility (HRF) payload on ISS was designed to try and prevent some of these lack of commonality issues by creating an HRF design guidelines document and a common software shell application. HRF is an ISS payload rack that supports many different studies using many different types of software built by different organizations. The guidelines document helped ensure that basic styles and layouts of all software packages were consistent. The common software shell application tied them all together, providing a common and consistent entry point to run the studies, and receiving high praise from the ISS crew.

The Space Shuttle Crew Reports reveal instances in which the design of display hardware was inadequate for the environmental conditions encountered. For example, on one mission, crewmembers reported that the sun shining through the overhead windows and the sun's reflection off the Earth through those windows resulted in the flight deck monitors being almost totally washed out and unusable. The small light shades, designed to prevent this occurrence, were insufficient. The monitors were used to maintain the proper corridors during rendezvous, determining drift and drift rates, and enable target reading to determine if a fly-out is required. A crewmember was required to constantly hold a large cue card over the monitors to block the reflection and improve legibility of the interface.

During one Space Shuttle approach and rendezvous with the ISS, crewmembers reported that the display on Monitor 1 was dim. They set the brightness levels to "full" and the contrast knob to a position to give them an optimal view of the target from the centerline camera. When the Space Shuttle was approximately 5 feet from contact, a shadow appeared on the docking target displayed on Monitor 1. This caused the displayed image to become so dark that no alignment information was visible. Even with the brightness set to full, the display on this monitor was unusable. About 2 seconds later, the automatic iris function of the camera made the target in the display visible again. Although the target image could not be seen on the monitor when the shadow appeared, postflight analysis shows that the video image recorded directly from Monitor 1 was dim but visible when viewed on the V-10 recorder. Crewmembers on a different mission reported that the manual iris control on the display and control panel was not user-friendly. When a crewmember attempted to manually close the iris, the video brightness "actually bloomed". After crewmembers put the camera back into "auto", the video returned to the previous picture. During the postflight debriefing, it was pointed out that manual iris control also reverts shutter speed and gain to default settings. To optimize the video, all 3 must be set together. However, shutter speed and gain could only be commanded through the display. This problem was thus clearly attributable to poor user interface design.

The ISS crewmembers primarily use laptops; the laptops are used regularly and dispersed throughout the ISS. The use of keyboards and pointing devices require special consideration in microgravity. Some crewmembers secure themselves by bungee to steady and restrain themselves while typing. In recent years, tablet use has increased on board ISS. Some crewmembers have indicated a preference for the use of a touchscreen interface for some activities because it allows then to work with more mobility than with a laptop secured to one location (FCI ISS crew comments database). These accounts highlight the additional considerations for display hardware during space travel.

Risk of Inadequate HSIA and Human, Automation and Robotics Integration Factors

Human-Automation Integration (HAI) in Spaceflight

From the beginning of the human spaceflight program, design of HAI has had a significant impact on mission success. For instance, during Apollo 10, poorly communicated information about actions and state changes of both automation and crew increased risk (Shayler, 2000). Specifically, at the end of the second pass over of the Lunar Landing site 2, the Apollo 10 crewmembers were preparing to separate the ascent and descent spacecraft components of the Lunar Module and return to the orbiting Command Module when the guidance and navigation system mode was inadvertently changed by one of the crewmembers. A couple of seconds later, the other crewmember reached up, without looking, and changed the mode of the guidance system, canceling the change made by the first crewmember. As a result, the Lunar Module, *Snoopy*, began firing its thrusters in all axes, pushing the gyroscopes into gimbal lock and rendering the navigation system useless until it was reset. The crewmembers then toggled the navigation system switch again and, although it was now put into the mode it should have been in to start with, the situation was made worse. At this point the crew overrode the computer and took manual control. The incident lasted about 15 seconds, during which *Snoopy* made 8 complete rolls. Had the crewmembers not regained control within another 2 seconds, it could have been too late to avoid impact with the moon. This close call was labeled “human error” (Orloff, 2000), however, nowadays, this incident is an example of state confusion, a negative HAI result due to poor design.

As human spaceflight has evolved, so have the technologies that enable continuous presence in LEO, and HAI design has had to support a range of LOA. Two main reasons exist for supporting a range: (1) astronauts are often required to be “backups” when automation fails and (2) requirements for spacecraft automation change during development, imposing different operational requirements on astronauts. If decided early in the system design process, HAI implementation can accommodate multiple uses. For instance, current space operations depend on autonomous rendezvous and docking of the Soyuz spacecraft to the ISS. HAI design must include ways for crew to monitor this dynamic, critical event but also provide manual take-over during unexpected anomalies (e.g., during Soyuz Transportation Modified Anthropometric [TMA]-5 mission [Energia, 2004] and Soyuz TMA-14 mission [Zak, 2009]). Additionally, although the initial Soyuz docking is automated, its crew manually controls relocation of the Soyuz from one docking port to another (NASA, 2012).

Not appropriately accommodating for multiple LOA led to a near LOC during the Apollo-Soyuz program (Lilley, 2016; Sgobba & Kanki, 2018). During descent and reentry, the Apollo Command Module could automatically deploy and activate the necessary elements for the Earth-Landing System, or the crew could manually trigger the required pyrotechnics and automatic sequencer for the Earth-Landing System. Unfortunately, the pilot of the Command Module armed the pyrotechnics too late and missed the automatic sequencer. The main parachutes were finally manually deployed at 7,150 feet instead of automatically at 10,500 feet. This sequence of events caused nitrogen tetroxide gas to leak into the cabin, irritating the crew's skin and eyes. The ensuing coughs impeded communications with ground control. Upon splashdown, the pilot was unconscious. The exposure to toxic gas required a 2-week hospital stay for the 2 crewmembers. Although a variety of contributing factors led to this incident, it shows that human-automation interactions require careful consideration with regards to function allocation and human performance under time-pressured conditions.

Spacecraft that have been operational for more than a decade still have critical system failures, for instance the environmental control and life support system has had critical issues during ISS's lifetime. Wu and Vera (2019) report that for the ISS, 67 high level IFI have been reported since 2000. An IFI is a record of an unanticipated off-nominal condition for the spacecraft system and its resolution. Although most of these IFIs were recorded early in the operational history of ISS (between 2000 – 2006), an average of one is still reported each year. These IFI require significant time for ground personnel to diagnoses, troubleshoot, and resolve.

Experience has shown that unexpected software anomalies are also likely during spaceflight missions. Mishaps attributed to safety critical software range from failed mission objectives (e.g., failed autonomous rendezvous satellite demonstration in 2005) to significantly reduced mission performance (e.g., 5-month long recovery of science from communication loss) (NASA Safety Center, 2012). The Space Shuttle, even after decades of operations, had in-flight software anomalies (NASA Safety Center, 2009). Avoiding automation brittleness extends beyond having "bug-free" in-flight software; it also requires consideration of how people interface and interact with the automation. During STS-32, the Space Shuttle's flight control system worked as expected, yet loss of attitude control (a safety close call) was experienced because of the uplink of an erroneous state vector (Ellis, 2000; Shayler, 2000; S&MA Flight Safety Office, 2012). Although automation failures are not predictable, HAI design should consider automation brittleness.

Evolution of operational concepts or creeping requirements during the system design process can affect HAI, as evidenced by the Mir-Progress collision. In 1997, the Russian spacecraft

Progress 234 collided with the Mir space station, causing the Mir pressure hull to rupture, and nearly causing the space station to be abandoned (Ellis, 2000; Shayler, 2000; see Figure 24). A contributing factor to this accident was the poorly designed operations concept and insufficient needs analysis, i.e., inadequate HAI design. The original operations concept was that Progress would dock automatically with the Mir, and because the crew responsibilities were primarily to cancel the approach if necessary, few displays were provided. Crewmembers were *later* tasked with performing the docking manually, after a decision not to install automatic docking systems for every Progress spacecraft. This change, however, was not followed up with an assessment of integrated human-system performance and, thus, neglected to consider the types of information crewmembers would need to remotely control another vehicle. Consequently, no additional displays were provided.

Figure 24. Spektr Module Showing the Damaged Radiator and Solar Array on Mir Space Station

The Space Shuttle Program provided another example of operational concepts that evolved and new requirements that were imposed on astronauts to “fill in” when desired automation was not implemented. The Space Shuttle was supposed to complete the *re-entry* and *landing phases* of flight automatically; however, certification of automated final landing was challenging, and eventually it was only certified as an “emergency system”(O’Connor, 2006). As a result, all Space Shuttle flights *landed* manually (Mindell, 2006), whereas, *re-entry* phase was automated and only one Space Shuttle flight (second Columbia flight) was ever executed manually (Mindell, 2006). Late system modifications that includes additional command and control capabilities for the crew resulted in *costly changes* in the Space Shuttle Program (McWhorter et al., 2010).

Human-Robotics Integration (HRI) in Spaceflight

NASA's experience with in-space robots includes robotic arms, rovers, in-flight free-flyers, and anthropomorphic devices. The majority of NASA's robotic experience is with large robotic manipulators on the Space Shuttle and the ISS. Crewmembers teleoperated these devices from the Space Shuttle and continue to do so on ISS. Ground flight controllers also teleoperate the robotic arms from Earth. Technology demonstrations have included ground operators sending commands to Robonaut2 (Joyce at al., 2015) (Diftler et al., 2015), a highly dexterous robot, and to the Smart Synchronized Position Hold, Engage, Reorient, Experimental Satellites (SPHERES) (Fong et al., 2013) , a free-flyer within the ISS. Crewmembers have experimented with in-flight teleoperation of a free-flyer, EVA Robotic Camera Sprint robot (Pederson et al., 2003), and a surface rover (Bualat et al., 2014).

Mars Surface Rovers

The Mars surface exploration robots, Sojourner, Phoenix, the Mars Exploration Rovers (MER), and Mars Science Laboratory, have been used extensively for over a decade but their operations are open loop, requiring ground operators to send sequences of commands. Real-time commands are not possible due to the communication transmission delays and the intermittent communication availability (due to satellite coverage) between Earth and Mars. Future robotic systems will be increasing more complex as they move from direct teleoperation to supervised teleoperation and autonomous control, with both spatial and temporal distances between operator and robotic agent. Employing human-robot teams will require a level of coordination not previously realized among teams of astronauts, robots, and mission controllers.

Risk is inherent to our lack of knowledge and experience of how best to design future HRI. However, many lessons have been learned from the years of the Space Shuttle, the ISS, and Mars robotic operations. MER trajectory planning is discovery driven, i.e., plans depend on previous sensor observations. This differs from orbiter or free-flyer missions, which are trajectory driven, i.e., observations are made while the vehicle follows a predetermined path relative to targets of interest. MER operations are constrained primarily by rover capabilities and a mission operation strategy that must accommodate 6- to 44-minute round-trip (depending on Earth-Mars distance) communication latencies plus the absence of a continuous communications link. These constraints mandate a multi-step human interaction process for the rover(s) to execute activities on the surface of Mars. Beginning with receipt of a downlink from Mars and assessment of its content, science activities are planned, refined, validated, reviewed, and then approved. This results in the generation of a command sequences that must be resolved and integrated, validated, and reviewed before commands are transmitted to the spacecraft. Some of the generalizable HRI lessons learned from the MERs include the following (Fong et al., 2013):

- High-fidelity resource modeling is essential because resources are tightly constrained.
- Day-long execution sequences are effective whereas they would not be in the absence of communication delays.
- Strict enforcement of the tactical timeline to reconcile conflicting science objectives enables mission success.
- Using onboard autonomy to travel beyond the range of data available to mission operations enables longer-distance traverses.
- Because a Mars-time schedule was not sustainable long term by mission operations, a modified Earth-time schedule that still produced an acceptable rate of science return was adopted.

McCurdy (2009) describes the iterative process that evolved the tools used by the Mars robot operators to plan and execute commands. After tactical processes were implemented, they found that they were not able to support the deadlines that the tasks demanded. Planning tools were developed to meet the necessary timelines, but these too suffered inefficiencies (Norris et al., 2005). After applying human-centered principles, several interface issues were identified. McCurdy (2009) found that many of the tools designed by different groups for differing purposes were inconsistent and non-cohesive, leading to steep learning curves as well as performance issues. Additionally, the primary timeline tool was based on outdated legacy software and was developed with a different user population in mind. This, in turn, depended on the operators' ability to make or miss an uplink deadline. Through process improvements and tools, the operations schedule shifted from 7 to 5 days a week. As a result of this analysis, McCurdy and his team were able to design and develop new tools for the next Mars mission, Phoenix Mars Lander, and further improved the Mars Science Laboratory.

Robotics in Low-Earth Orbit

Optimized robotics operations are essential for spaceflight operations, including ISS maintenance and Extravehicular Robotics (EVR). For overview, see Chang & Marquez (2018). According to data from the FCI ISS crew comments database, this includes the proper set up and configuration of the onboard Robotic Workstation (RWS) and subsequent actuation and execution of arm operations. The RWS for the SSRMS consists of a laptop computer, translational and rotational hand controllers, 3 video monitors and a display and control panel. Robotics workstations on board the ISS are currently located in the U.S. Lab, Cupola, and the Japanese Pressurized Module. Although the workstation can be operated by a single crewmember, in practice, RWS activities are conducted with 2 crewmembers. One crewmember acts as the primary controller, and the second crewmember controls cameras, tracks procedure steps, and confirms the direction of motion. Crewmembers report that this allows them to increase their SA and maintain a system of checks and balances, although the second operator does not necessarily have to be another on-orbit crewmember, and ground personnel routinely support EVA and EVR as the second robotics operator. In addition to the base equipment provided at the RWS, operators rely on additional camera views accessible on laptop computers positioned near the RWS and visual cues, such as targets overlaid onto video feeds. Ground personnel can download up to 6 distinct video feeds and can also use overlays to perform contact operations (e.g., grappling and grasping).

During postflight debriefs, ISS crewmembers have consistently praised onboard robotics training, including the simulation fidelity and the appropriate level of difficulty of training sessions. Crewmembers can also refresh their robotics skills on board using training tools such as the ROBoT-r (see Figure 25, Figure 26). Furthermore, ground personnel are still an integral part of onboard training systems, verbally interjecting real-time. Although this arrangement is sufficient for ISS, Mars-bound astronauts will need onboard robotics training for tasks they have never performed on Earth, and communication with trainers will be subject to delays. Based on previous crew comments, the fidelity of future robotic training simulators may heavily influence crew efficiency and effectiveness.

Figure 25. NASA Astronaut, Karen Nyberg (Increment 36/37) Performing Work at the Robotics Workstation in the ISS Cupola

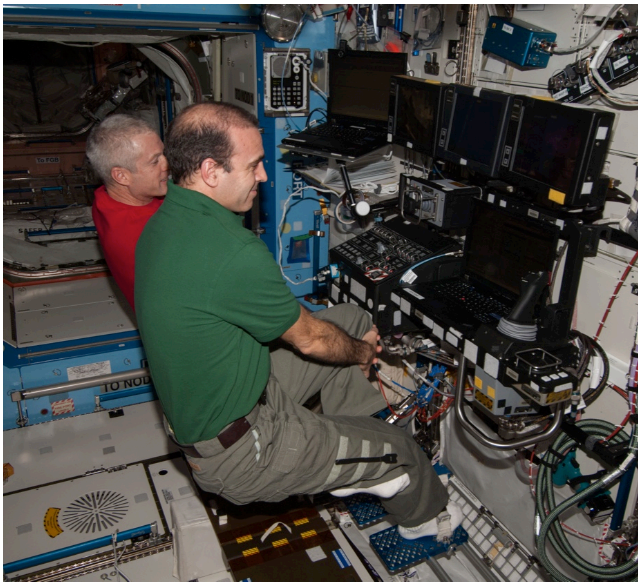


Figure 26. NASA Astronauts Rick Mastracchio and Steve Swanson (ISS Increment 39) Working ISS RWS in the U.S. Lab, with Hand Controllers, Display and Control Panel, and Monitors in View

Figure 27. Completing an EVA Activity using the Robotic Arm with a Crewmember on the End Effector

Manipulation of the ISS robotics arm is essential for EVA (Figure 27) and robotics operations are critical for berthing visiting vehicles. The SSRMS is used to capture visiting supply spacecraft such as SpaceX Dragon or the Japanese H-II Transfer Vehicle, maneuver them up to a docking port, and berth them. Additionally, the visiting vehicle is undocked and moved by the SSRMS to a safe distance and trajectory from ISS for its return to Earth.

The berthing of visiting vehicles is a procedure adapted from previously existing robotics tasks (Figure 28). During vehicle berthing, the crew arranges a robotics workstation within the Cupola and the whole operation is coordinated between ground teams and crewmembers. The crew performs the capture and release of the free-flying vehicles, and the ground operators maneuver the vehicle from the capture position to berthing, as well as the unberthing and maneuvering to the release position. This procedure required careful evaluation of lessons learned for systems, procedures, and operations collected over the life of the ISS.

Figure 28. Visiting Vehicle Dragon Captured on the SSRMS End Effector (Increment 40)

The adaptation of existing robotics operations to support new robotic tasks provides unique HARI spaceflight evidence on the effectiveness of current HRP guidelines and standards for HARI design (NASA, 2014) in an operational context. Crewmember comments during ISS debriefs indicate that NASA may benefit by enhancing HARI guidelines for the design of robotic workstations and out-the-window views for robotic operations. Although visiting vehicle robotic operations are conducted at robotic workstation in the Cupola, which affords crew a direct view of the spacecraft and the SSRMS, SSRMS manipulation for EVA-related robotics operations employs the robotic workstation in the U.S. Lab module. Crewmembers have commented that, initially, limited guidance was provided on how best to set up the robotic workstation in the Cupola. Additionally, many comments have been provided regarding the out-the-window lighting conditions (i.e., too bright or too dark) and their potential effect on performance (e.g., ability to assess out-of-window views such as relative size or changing rates of motion). Ground teams immediately addressed these crew comments to better support visiting vehicle robotic operations on the ISS. Moreover, these lessons learned are essential to consider when developing guidelines and standards for future robotics operations because future missions may present similar challenges.

Given the already strained robotics resources and the large number of robotic and EVA operations required for ISS, ground control personnel could be used to off-load on-orbit crew workload and maximize the efficiency of ISS operations (Coleshill et al., 2009). In 2005 during STS-122/ISS 10A, these efficiencies were finally realized when the majority of pre- and post-operation configurations for SSRMS were allocated to ground control (Aziz, 2010). Adjustments had to be made to reduce the constraints on the step size of arm motions to allow greater movement distances in less time. These adjustments were originally implemented by safety personnel to reduce the duration of each maneuver, thereby limiting the number of unplanned lost communication signals between the ground and the ISS. Task analyses determined the best allocation of operations between the on-orbit crew and ground personnel. This type of crew-ground coordination will be critical as control of remotely located robotics becomes more commonplace in an environment with increased time delay and loss-of-signal events.

At present, ground operators and on-orbit crew share many of the EVA and EVR responsibilities (Webb et al., 2009). Ground control is used to the greatest extent possible to reduce EVA time and increase the probability of completing EVA goals. For future planetary surface operations, this could involve an operator in a habitat controlling an external robot to maximize the efficiency of surface EVAs. Currently, barriers to extensive ground control exist because mission controllers restrict the allowable types and methods of operation. Additionally, the robotic systems and their interfaces were designed for on-orbit use and are not tailored to the needs and limitations of a remote operator. As a result, ground controllers spend significant amount of time preparing for ground-led operations, which includes many days validating and verifying that the robotic arm commands will successfully execute in space and operate within the prescribed safety limits.

As seen in Figure 29, the Special Purpose Dexterous Manipulator (SPDM), Dextre, is a 2-armed manipulator mounted on the SSRMS as part of the Mobile Servicing System. SPDM is designed to perform dexterous robotic maintenance, payload servicing, and other miscellaneous tasks. Before the system was launched in 2008, it was determined that the crewmembers would require extensive training to use the SPDM, and operating the SPDM within the accepted constraints was beyond the available crew resources (Aziz, 2013). A NASA Technical Interchange Meeting that convened in 2002 determined that the prohibitively long timelines for the SPDM tasks were due in part to the design, which requires one arm for stabilization and a single arm for conducting tasks (as opposed to 2-arm operations). Initially it was determined that the precise positioning required to perform dexterous operations required one arm to provide stabilization at the worksite while the second arm performed the dexterous task, however, early on-orbit tests showed that this was not the case. Tasks are now designed to use both SPDM arms to more efficiently complete operations.

Aziz (2013) describes the elaborate process of validating ground-based telerobotics control of the SPDM. The robotic arm was originally designed to be operated by ISS crewmembers, and as a result, its operational functions did not meet safety constraints imposed on ground-controlled telerobotics operations: "Human-in-the-loop control modes were not approved by the safety community for use during ground-based Dextre or Canadarm 2 operations." (pg. 3; Aziz, 2013). Ground controllers must use deterministic preplanned maneuvers and end-effector sequences to control the robotic arm, whereas astronauts can directly manipulate the robotic arm. Consequently, major software updates were required, resulting in significant changes in the types of commands the operator could input. This development is evidence that even in most common space robotic systems (i.e. a robotic arm), requirements often evolve without evaluating the costs associated with HARI.

Almost all removal and replacement tasks require a large number of procedural steps and robotic tool manipulation (Caron, 2004). Procedures are entirely preplanned and approved, leaving little to no ability to respond in real-time to anomalies or contingencies. This affected an SPDM activity in December 2010 wherein an anomaly occurred and operations had to be halted for over a day and re-planned. Several enhancements such as auto-sequencing for joint angles and targeting overlays have been implemented to help mitigate these issues.

Figure 29. In June 2008, Dextre was Moved Atop the Destiny Laboratory Module of the ISS, Before Deployment of Japan's Kibo Pressurized Science Laboratory

Experience has shown that the SSRMS can be operated by either the ISS crew or by ground control (Fong et al., 2013). The latter scenario introduces additional complexity due to 3-10 s latency of time-varying command and telemetry. To mitigate the impacts of these delays and of data dropouts, ground control of SSRMS is limited to execution of automatic control sequences, and even though real-time input is not used in automatic operation modes, ground must still maintain vigilance and SA during manipulator moves. Moreover, ground-based operation is restricted to predicted periods of reliable communications between the ISS and the MCC.

Among SSRMS lessons learned are the following (Fong et al., 2013):

- No single control mode is sufficient for all operations.
- Maintaining SA requires multiple options for handling coordinate system reference frames and graphics tools.
- Off-loading routine tasks to ground is beneficial because it frees up valuable crew time.

The SPDM lessons learned include (Currie & Peacock, 2002; Fong et al., 2013):

- Standardizing the mechanical interfaces decreases time and costs associated with various design, planning, and execution phases.
- Task completion times extend because operators rely on visual features under extreme lighting variations.
- Task times might be reduced through better graphical interfaces and automated function.
- SPDM tasks completed by ground teams frees up valuable crew time.

One human-factors-centered objective in the design of SSRMS and SPDM is common display-panel graphical user interfaces (GUI) (e.g., symbols and color coding) and common manual devices for the 2 manipulator's control stations and the ground and ISS control stations (Currie & Peacock, 2002; Fong et al., 2013). To maintain uniformity of interfaces with the SRMS a 3-DOF rotational translational hand controllers was selected rather than a controller combining all 6 DOF in a single integrated 6-DOF unit. Common manual interfaces reduce training needs and maintain already formed "positive" habits that could reduce the potential for inadvertent or errant manipulator commands during manual control (Currie & Peacock, 2002). Furthermore, Currie and Peacock (2002) and Fong et al. (2013) contend that control stations should allow for emergency operator intervention through hardware rather than software interfaces because of the potential hazard introduced by having to negotiate through several layers of software to stop of the manipulator. This lesson learned is currently applied to the RWS, which includes hard switches for critical commands such as applying or removing SSRMS brakes.

SA in space is paramount for robotics operations. Several crewmembers have commented during their ISS debriefings that more than one astronaut is required to perform robotics tasks to improve and maintain SA. Moreover, SA between ground personnel and crewmembers is critical, particularly when sharing robotics operations roles and responsibilities. Crewmembers have commented that robotics task hand-offs between the ground team and astronauts are smooth and observed that SA is essential for this transfer of responsibilities. Crewmembers have indicated that it is commonplace for astronauts to wait for ground to confirm or acknowledge next actions, they consult regularly with the ground team during robotics operations. As crew autonomy increases during future LDEMs, maintaining SA during space robotics operations may be more challenging. Crews will have to contend with communication time-delays, less interaction, or potentially no dialogue with ground team during time-critical events. Future HARI design must include mitigations that support the crew under such communication constraints, and in turn, provide as much feedback to ground controllers who will have to follow-along without much opportunity for intervention.

One particular example of a time-critical event is one that occurred in 2013 (Svitak, 2013) when a visiting vehicle executed an abort departure within minutes of being released from the SSRMS. Within that short timeframe, the crew commanded release, retracted the robotic arm, and assessed the impending departure of the un-crewed spacecraft. Operational procedures placed the visiting vehicle out of harm's way of the ISS and the robotic arm. However, an unexpected mechanical interaction between the robotic arm and the spacecraft introduced undesirable motion that caused the spacecraft to approach the ISS. Crewmembers were ready to signal the abort pending ground confirmation, but the visiting vehicle correctly executed the abort maneuver before any human intervened.

This close call demonstrates that current spaceflight operations can integrate human capabilities with highly autonomous systems and robotic systems. Crewmembers have expressed confidence in the robotic operations and the spacecraft's automated onboard flight system. Additionally, crewmembers have commented on the importance of evaluating all possible mitigation strategies for tasks that takes less than 2 minutes to execute (e.g., sending the abort signal or moving the robotic arm further out of reach) and of receiving confirmation from ground. Future LDEMs will have to adapt spaceflight operations for autonomous performance of robotic tasks by providing effective user interfaces, cognitive aids, and training capabilities.

Advanced Robotics Demonstrations in Space

Arriving aboard the ISS on February 2011, Robonaut 2 (R2) is a humanoid robot that consists of a 3-DOF head, 1-DOF torso, 2 7-DOF arms, and 2 12-DOF hands. R2's "human-like" dexterous hands provide the capability to use the same tools, handholds, and interfaces as astronauts. R2 is envisioned to be a teammate to assist astronauts with repair and maintenance tasks. Ground controllers and ISS crewmembers can remotely operate R2 through a customized R2 GUI and a head-mounted VR system. Because R2 is a research and development project, the software and control interfaces are still under development, Figure 30.

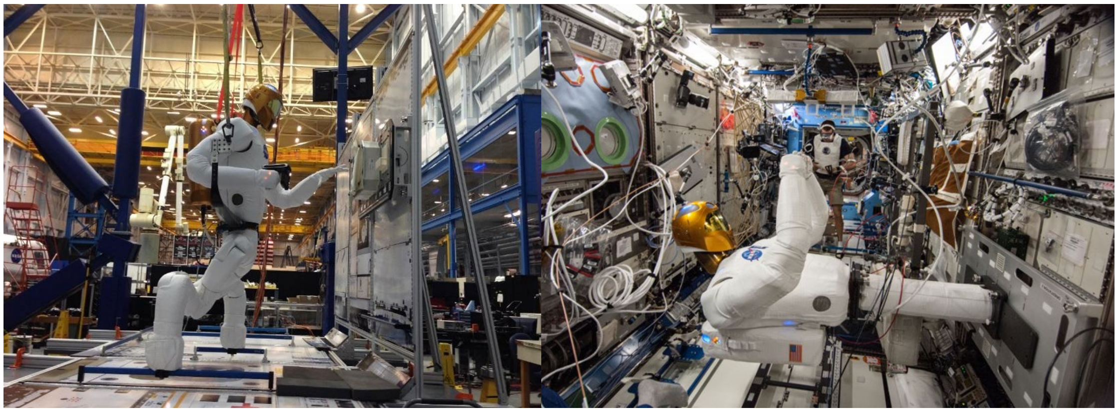


Figure 30. A Demonstration of R2 on Ground (Left) and Robonaut On Board ISS (Right)

The R2 ground control team includes an operator, a task monitor, and a lead. The operator monitors the R2 GUI and the task monitor oversees the mission and task timelines, assisting the operator during anomalies. The lead oversees the entire R2 operation and communicates with the rest of the ISS MCC team (Fong et al., 2013). Although the R2 is a promising new robotic system to support space operations, no published data currently exists regarding its performance.

SPHERES is a high-fidelity test bed designed for developing and maturing algorithms for distributed satellite system concepts, including autonomous rendezvous and docking maneuvers. The SPHERES program began as a design course in the Space System Laboratory at the Massachusetts Institute of Technology, and developed over the years into a permanent robotic experiment on the ISS (Stoll et al., 2012). The first SPHERES experiment on the ISS was in October 2003. SPHERES consists of 3 free flying spherical satellites that are 0.2m in diameter and weigh 3.5kg. Over the years, SPHERES as served as a platform to conduct various ISS experiments (Interact, InSPIRE, VERTIGO, Slosh, STELLA).

The crew have limited interaction with SPHERES and are only required for experiment setup and cleanup. The SPHERES operations team includes the ops lead, ops support, telemetry, video and voice recorder, activity logger, and PI. The ops lead monitors the progress of procedures and communicates with the crew during the sessions. The ops support monitors the mission and task timeline. The telemetry position monitors the telemetry data from SPHERES. The video and voice recorder ensure that video and audio data are being recorded. The activity logger logs activity notes. The PI also monitors the progress of procedures and aids in troubleshooting anomalies. A NASA (unpublished) internal study on human-robot user experience revealed that for SPHERES, pragmatic qualities are of higher importance than to hedonic qualities.

As of 2019, other free-flyer robots have been deployed to the ISS: Crew Interactive MObile CompanioN, Internal Camera Ball, and Astrobee. MObile CompanioN is an AI-based assistant funded by the German Aerospace Center, Airbus, and powered by IBM’s Watson. Funded by the Japanese Space Agency, the Japanese Experiment Module Internal Camera Ball aims to provide photo-video documentation in flight. The goal of NASA’s Astrobee is to reduce astronaut time by conducting routine tasks. Data has yet to be published on these space demonstrations. Astrobee will have a guest scientist program, like that of SPHERES, that will allow researchers to study human-robotic interaction in space.

Risk of Inadequate HSIA and Training Factors

Detailed studies of the effectiveness of spaceflight training are rare. Most data are documented in post-mission crew debriefs, in interviews with training SMEs, in anecdotal reports from oral histories, or within incident reports or accident investigations. Thus, most of the spaceflight evidence presented for training contributing factors will be either case studies (Evidence Category III) or expert opinion (Evidence Category IV).

The primary sources of the expert opinion data are:

- **The FCI ISS Crew Comments Database.** Beginning with the return of the first Expedition Crew to the ISS in 2001 and continuing on today, NASA’s ISS Program has conducted structured debrief interviews to document crew observations and concerns related to working and living on board the ISS. The debriefs cover approximately 2 dozen topics such as vehicle systems, food, habitability and human factors, and training. The FCI team at NASA collects and analyzes human-related data from each debrief, deidentifies the data to protect crew privacy, and stores the data in the FCI ISS crew comments database.
- **The FOD Stakeholder Concerns and Top Priorities Workshop Report.** In 2018, HRP Training Risk and Team Risk researchers conducted a series of 3 half-day face-to-face workshop interviews with FOD personnel supported by 2 crewmember representing the astronaut office, the Acting Branch Chief of the FOD Training Branch, a Chief Training Officer also from the Training Branch, and a Senior Systems Instructor (Dempsey et al., 2018). In preparation for the FOD workshop interviews, an HRP researcher met with 2 crewmembers the week prior; one of these crewmembers also attended the 3 half-day meetings. The preparatory meeting and the first 2 half-day workshops consisted of unstructured interviews to document training stakeholder concerns. The third half-day workshop consisted of documenting FOD’s top priorities among their concerns.
- **The ISS Best Practices Interviews.** In 2016 and 2017, HRP Training Risk researchers conducted a series of structured interviews with current and former training personnel to document training best practices during the Space Shuttle and ISS Programs (Dempsey et al., 2018). The interviewees included Space Shuttle and ISS flight controllers and instructors, Station Training Leads, and training managers. The interview questions were structured to cover topics including training design and redesign efforts, training evaluations, and learning and performance support.

- **The ASCAN Working Group Poll.** In 2017, an HRP training risk researcher conducted a poll of the ASCAN working group, the team of instructors assigned to train NASA's 2017 ASCAN class. The poll consisted of a simple question asking each respondent whether they expected that the current program of ISS training for their system or discipline would be effective for training autonomous or semi-autonomous operations for future LDEMs and asking why or why not they expected this.

Foundational Principles of Training

Spaceflight Evidence: Retention

The current program of training does not support retention across the large number of duties necessary for crewmembers to operate autonomously or semi-autonomously from MCC during LDEMs. During a FOD training stakeholders' workshop supported by astronauts and training personnel, training retention was named the "Priority 1" training concern (Dempsey et al., 2018). Workshop attendees stated that lack of training retention during spaceflight is a consistent concern expressed by training leads, instructors, astronauts, and training management, and attendees also stated that crewmembers unofficially report memory and performance issues on board. However, to date, NASA has not conducted controlled studies of onboard retention, so no objective data exists quantifying the retention duration and domains most severely impacted.

Nonetheless, an abundance of evidence exists from crewmembers, training SMEs, and spaceflight case studies that shows that retention of training is an issue. During postflight debriefs, ISS crewmembers consistently comment that they did not remember details of ground training, that they did not recall being trained on specific hardware or payloads, that they did not remember specific briefings, and/or they had a hard time remembering training that was, at times, conducted 2 years before their mission (FCI ISS crew comments database, n.d.). Landon et al. (2018) described the challenges of retaining both technical and teamwork training, including communication and team cohesion skills, stating, "Unfortunately, current technical and teamwork skills training for astronauts has not been evaluated for effectiveness, and teamwork skills suffer from similar [retention] problems related to long lag times between preflight training and mission deployment." (p. 569) Finally, a redesign of the ISS training flow—the FOD training reduction effort—was undertaken circa 2015. The training flow duration was reduced from ~30 months to ~24 months by reducing training content, addressing procedure issues, and reducing refresher training. The FOD undertook the effort to reduce the flow because of, among other things, "knowledge retention issues" (Sonoda, 2016). Although reducing the training content in an effort to address retention in the remaining flow is a solution that works for low-Earth orbit where real-time support is readily available to augment content not trained or not retained, such a solution is not feasible for missions to Mars, indicating a need for developing more effective preflight training.

Figure 31. A Mir Solar Array with Damage from a Collision with an Incoming Progress Spacecraft on June 25, 1997

This lack of training retention has existed during many spaceflight programs and led to near-devastating consequences during Russia's Mir program. The collision on June 25, 1997 of the arriving Russian spacecraft Progress 234 and the Spektr module of the Mir space station during an attempt by a cosmonaut to manually dock the incoming Progress, caused Spektr's pressure hull to rupture, leaving the crew in a spinning space station without power or computer control, and nearly causing the Mir to be abandoned (Ellis, 2000; Shayler, 2000) (Figure 31). Multiple factors contributed to the Mir-Progress collision (Ellis, 2000) one of which was the training-performance interval. Although performance declines are associated with increasing intervals between training and performance (see, e.g., Winfred, et al. 1998), the cosmonaut performing the manual docking attempt had last received formal training on the teleoperated rendezvous control system 4 months before the docking incident and was not provided refresher training prior to task execution, "leaving his docking skills degraded" (NASA, 2010).

Without a good understanding of the training retention curve, especially for mission critical tasks, and without effective CMs for a lack of training retention, there is a possibility of adverse outcomes in the performance of critical mission tasks.

The space environment (e.g. isolation & confinement, radiation, elevated CO_2, and microgravity-induced fluid shifts) imposes significant physiological, psycho-social, and cognitive loads on the crewmembers that is likely impact their performance (Dempsey et al., 2018; Slack et al., 2016). To date, no systematic data has been collected to understand the effects of such loads on crewmembers' ability to retain trained knowledge and skills, and to transfer such knowledge and skills to novel situations. Barshi et al. (2019) attempted to conduct the first such study to systematically assess the effects of long-duration spaceflight on training retention and transfer.

Because current theories of retention and transfer are based on results obtained in university laboratories using undergraduate students as research participants, and because crew time in space is very expensive, Barshi et al. (2019) designed their study to compare the performance of 4 groups of subjects: crewmembers in space, crewmembers on the ground, crew-like subjects, and university undergraduate students. The 3 groups of ground subjects completed an identical 16-month protocol that included 2 experimental tasks: a standard data entry task with variants (for motor control and information processing) and a continuous memory-updating paradigm using both retrospective memory (remembering something from the past) and prospective memory (remembering to do something, in this case something different, in the future). The study demonstrated that crewmembers' performance under cognitive load cannot be adequately predicted from the performance of university undergraduate students. Unfortunately, this study was ended before assessment of the crewmembers in space. Therefore, the extent to which crewmembers' cognitive performance in space can be predicted from the performance of crewmembers on the ground is unknown, and, to-date, no systematic study of in-mission retention of preflight training has been performed.

Research is needed to determine the retention interval for the skills trained to future crewmembers of LDEMs, especially for mission critical tasks, and it is important that the research subjects in such studies are well-matched to crewmembers. The retention intervals determined from this research can then be used to design a program of training across the entire training continuum.

Spaceflight Evidence: Transfer

Adverse performance outcomes can result when the crew are unable to adequately respond to unanticipated critical malfunctions. The current program of training does not support the transfer of trained content to novel tasks encountered in the operational environment, including critical malfunction scenarios.

Given that the consequences of a major power failure on the ISS can be severe, crew are provided extensive training on the electrical power system and the planned response to a critical electric power loss. Nonetheless, during 60% of the critical power failures on ISS crewmembers responded incorrectly despite having been provided preflight training on the correct response. (To date, there have been 5 critical power losses on ISS.) An evaluation of the training effectiveness showed that the failure scenarios encountered on ISS were not identical to the training failure scenarios before flight. Training updates made subsequent to these adverse performance outcomes included adding additional variations to the trained scenarios. In 2018, crewmembers requested onboard proficiency training for this skill, and that training has been developed. None of these incidents resulted in catastrophic loss because MCC provided real-time ground support to mitigate the loss of training retention, detected safety critical procedure errors, and supported the crew in responding to the failure. This safety-critical real-time ground support will not be available during future LDEMs due to the communication delays.

As this evidence shows, without an understanding of how to design a program of training for effective transfer of training, there is a possibility of adverse outcomes in the performance of critical mission tasks.

Spaceflight Evidence: Simulator Fidelity and Training for Expertise with Automation

NASA's human spaceflight program embraced simulation training soon after the Agency's inception. Even facilities that were initially employed to test human tolerance for extreme environments (e.g., high-G, low atmospheric pressure) were soon modified to incorporate flight-relevant tasks (Woodling, et al., 1973). However, although NASA has provided training simulators for each of its human spaceflight programs, if the appropriate level of training simulator fidelity is not provided, there is a possibility of adverse performance outcomes.

The most tragic example of this occurred on October 31, 2014, when the co-pilot of SS2 issued a single command to prematurely unlock the feather system resulting in fatal injuries to the co-pilot, serious injuries to the pilot, and structural breakup of the vehicle. The NTSB concluded that the probable cause of the accident was a single human error that resulted in a catastrophic hazard to the vehicle. Although Scaled Composites, the manufacturer of SS2, "identified 'uncommanded feather operation' as a catastrophic hazard during the boost phase, with the 'probable loss of aircraft' as the resulting effect on the vehicle and flight crew … [they] assumed that the flight crew would be properly trained through simulator sessions." (NTSB, 2014a, p. 41). However, the NTSB stated that, "Scaled Composites did not ensure that the accident pilots and other SS2 test pilots adequately understood the risks of unlocking the feather early." (p. 46). Furthermore, the NTSB specifically questioned both the physical fidelity and motion fidelity of the simulators provided to train the SS2 crew.

> "According to SS2 test pilots, pilots did not train in the simulator with flight suits, helmets, oxygen masks, parachutes, or gloves, which they would be expected to wear during actual flights in the SS2 vehicle. Because Scaled did not require test pilots to train in their flight gear, human factors limitations in the cockpit might not have been apparent until an actual flight. For example, during the accident flight, the pilot stated (before starting the L-10 checklist), "this headrest, [at] least the height I'm at, when I press up against it I kinda lose the bottom half of the [PFD] screen." The copilot responded, "I remember doing that on, on [PF01], I took my head, consciously took my head off the headrest." The information displayed on the lower portion of the PFD might not be needed during the boost phase of flight, but a pilot should not be surprised by this or similar conditions during an actual flight.
>
> Also, even though Scaled's SS2 simulator replicated the SS2 cockpit, some aspects of the SS2 operating environment were difficult to model in the fixed-base (no motion) simulator, including the high G forces and vibration during flight. As a result of the lack of motion in the SS2 simulator, pilots were unfamiliar with the vibration and loads to be expected during powered flight. Further, according to a Scaled project engineer, the force required to unlock the feather in the simulator was less than the force required to accomplish that action in the SS2 vehicle. (The forces were not quantified.) Thus, test pilots did not have a realistic expectation of the actual force needed to move the feather lock handle or the mechanics of operating the handle with the vibration and loads that

> would be occurring. If Scaled had a designated human factors expert on its staff, that person's expertise could possibly have been used to address this and other human factor limitations in simulator training, resulting in a training environment that would better prepare pilots for an actual flight." (NTSB, 2014a, p. 46)

The NTSB stated that human factors should be emphasized in simulator training "to reduce the possibility that human error during operations could lead to a catastrophic event." (NTSB, 2014a, p. 68)

A second incident that showed the consequences of a mismatch between the training simulator and the actual vehicle is the landing of the third Space Shuttle mission, STS-3. The final detailed test objective for the STS-3 was a demonstration of the autolanding software until just after slowing down under Mach 1.0 when the commander would take over manual control for the final flare, touchdown, and rollout. In an oral history of the mission, the commander described the challenges in safely landing the vehicle using the autolanding software. The flight software was not programmed in a way that mimicked the way a pilot would fly the vehicle, "Automatic control of wide swings in speedbrake position did not mimic typical manual pilot inputs of small corrections to maintain constant airspeed." (NASA, 2010, p. 24) The high speed of the landing caused the commander to fly to a short landing position. And, an instability in the longitudinal control software caused a late, unexpected flare-up just before the nose touched down that required the commander to execute a quick pitch-up with the rotational hand controller, twice. Although the commander was able to land the vehicle safely, none of these issues, which all occurred within the last few minutes of flight, were encountered in training simulations. A later investigation determined that the autolanding software in the SMS did not match the flight software on the real vehicle,

> "which was in total violation of long-standing requirements to incorporate flight software timely in the SMS. Auto approach training in both the fixed-base and motion-base simulators never exhibited the unacceptable speedbrake control experienced in the STS-3 mission. The software in both of those simulators modulated the speedbrakes in small increments, as a pilot would do, to control speed very precisely on the OGS [Outer Glide Slope]. That is, we were led to believe there were no deficiencies in flight software related to OGS speed control, so the actual flight behavior in this regime was a complete surprise to which we were required to react in real time; exactly what preflight training is conducted to avoid!" (p. 27)

An additional mismatch occurred between the physical fidelity of the simulator and actual vehicle in the stick control input needed to disengage the autolanding software. As described by Charlie Bolden, a then astronaut and later the NASA Administrator,

> "Everything seemed to be going well until just seconds before touchdown, when all of a sudden we saw the vehicle kinda pitch up and then kinda hard-nose touchdown. We found out that, just as Jack Lousma had trained to do, you need to move [the stick] an appreciable amount [to disengage the autoland]. We didn't realise that. The way he had

> trained was just to do a manual download with a stick. When he did that, he disengaged the roll axis on the shuttle, but he didn't disengage the pitch axis, so the computer was still flying the pitch, although he was flying the roll. Gordon Fullerton just happened to look at the eyebrow lights and he noticed that he was still in auto in pitch. He told Jack, and so Jack just kinda pulled back on the stick and it caused the vehicle to pitch up. Then he caught it and put it back down and he saved the vehicle." (Evans, 2011)

While this evidence shows the critical need for effective training simulators and simulations, NASA does not yet have standards for the design of such simulators nor for the level of simulation necessary to support skill acquisition, retention, and transfer. An internal NASA report on the level of simulation fidelity needed for a future HLS for lunar missions acknowledges the HSI challenges in developing simulator training standards,

> "The problem of how to characterize intentional human performance so as to design systems and training [simulators] to best support proficient performance is a long-standing one in Human Systems Integration (HSI). The natural tendency is to recommend the highest level of fidelity for training purposes, at least for tasks deemed critical:
>
> *"The most important tasks must be trained to the level of skill-based behavior and to the extent that the astronauts can apply effective stress responses, and this requires training in real, stressful environments. Thus, sufficient training cannot be provided only in cheaper, low-fidelity simulations or classroom environments."* (NRC, 2011, p. 94)
>
> But this prescription skirts the issues of how to define which tasks are important and what makes a training environment "real" and properly "stressful."" (Null et al, 2019, p. 27)

The approach the FAA takes in developing its Part 60 classification standards for aviation flight simulators is acknowledged to be subjective (standards are based on experienced pilot's recommendations) and requires that pilots be provided flight training hours in training flights (i.e., the actual aircraft in which the pilot seeks to be rated). Although the FAA approach can provide some guidance, NASA needs a more principled approach that links task requirements for new vehicles to simulation requirements. (There are no experienced astronauts on new vehicles to provide recommendations and spacecraft are not used for training purposes—astronauts must master all of the skills for piloting before they launch.) It will be important to establish objective training standards for future spacecraft based on the best principles of training and HSI, rather subjective judgment. Research on human performance is required to establish such standards.

Finally, it should be noted that while the risk evidence presented here relates to piloting skills, is important to provide crewmembers the appropriate level of simulator and simulation fidelity across all high-risk, critical tasks including those that require an understanding of the automation and intelligent systems and those that require use of advanced decision-making tools. For simulation training needs for team skills see the Evidence Report: Risk of Performance and Behavioral Health Decrements Due to Inadequate Cooperation, Coordination, Communication, and Psychosocial Adaptation within a Team (Landon, 2016). For simulation training needs for medical skills see the Evidence Report: Risk of Adverse Health Outcomes and Decrements in Performance due to In-Flight Medical Conditions (Antonsen et al., 2017).

Training Program Design

Spaceflight Evidence: Analysis: Duty and Task Allocation

As stated above, the expectation is that each crewmember of future LDEMs will be assigned different jobs with unique sets of duties and tasks. However, the training experts in the FOD have expressed concern about how the duties and tasks will be allocated both among crewmembers and between the crew and the ground, and they have expressed concerns about how to define the correct level or depth of training content across domains given these duty and task allocations (Dempsey et al., 2018). Research is needed to develop a systematic method for allocating duties and tasks from a learning perspective. Additionally, given that task coherence, team decision-making, and meaningful work affect team cohesion and behavioral health (Landon et al., 2016; Slack et al., 2016), any duty and task allocation method should take these factors into consideration as well.

Spaceflight Evidence: Design: The Training Flow Across the Training Continuum

Instructors of multiple disciplines recognize that current ISS training will not be effective for LDEMs. As noted in a 2017 poll of the ASCAN working group, of most concern is that current training assumes the real-time system expertise and performance support provided by the ISS flight control team and the real-time expertise provided by scientists on the ground supporting in-mission research. For example, the flight control team currently performs actions such as system activation and checkout, vehicle safing actions and verification, and malfunction responses (current ISS training "assumes the ground will take care of the majority of the problems"). Additionally, the flight control team understands procedure details at a much deeper level than the crew does, including the reason for the sequence of steps, constraints, impacts, parallel paths, and expected results, and the flight control team has a wealth of knowledge and other flight control team members (including Mission Evaluation Room personnel) available to them to solve challenging problems.

Given the communication delays associated with LDEMs, real-time ground support will no longer be possible, and the crew will need to be trained for both nominal operations and off-nominal responses during semiautonomous and autonomous mission phases. Although it is expected that the flight control team and ground-based scientists will continue to provide expertise on systems and payloads during semiautonomous mission phases, team coordination and processes will be affected by the communication delay, which is not adequately addressed during current training. For autonomous mission phases, the crew will require software, data, or tools to troubleshoot system failures on their own, and they will need to be trained to perform such troubleshooting. It is also assumed that future vehicles will have more automated systems; however, current ISS crew training does not address training concepts such as trust in automation. Additional concerns about current training include the level of scientific expertise that will be needed for a planetary mission and the type of training needed to support this such as intensive training in the field, details of sample collections, contamination control, and in-situ measurements (Dempsey et al., 2018).

Spaceflight Evidence: Design Methods

Crew training should be guided by principles of skill acquisition, retention, and transfer that have been validated in studies of the psychology of learning. Dempsey and Barshi (2020) imagined a spaceflight training program that employs methods based on validated principles or learning (including variability of practice, easy-to-difficult ordering, strategic use of knowledge, spacing, and focus of attention). However, to date there have been no studies on how effective these or other methods are for the tasks required for future spaceflight missions. Given the retention issues with NASA's current program of training, methods beyond simple practice need to be investigated.

Spaceflight Evidence: Evaluation Methods

The spaceflight incidents and the expert opinions described in this report provide evidence that the current system of siloed training coupled with rehearsal or practice (though rarely any other training methods) and evaluations by system-specific mastery lessons do not produce the level of autonomy nor the durability of skills astronauts will require for the challenges of future LDEMs. In fact, during a training workshop, the FOD acknowledged that the masteries allow the crew to demonstrate knowledge of different systems such as electrical power and thermal control, but these masteries are not at all representative of the job (Dempsey et al., 2018). Nonetheless, as noted above, there are no controlled research studies of the spaceflight training program and no evaluation measures or methods. Research is needed to define evaluation measures and methods for future missions and then to validate such measures as evidenced by in-mission performance.

Performance Support

During a FOD training stakeholders' workshop, astronauts and training personnel expressed concerns that the current performance support tools will not support autonomous or semi-autonomous mission operations, and acknowledged that procedures are sometimes poorly written thus requiring MCC support to properly execute (Dempsey et al., 2018). Poorly written procedures do not necessarily indicate a lack of diligence on the part of the flight control team. For example, if vehicle configuration documentation differs from the actual vehicle layout, procedures based on the documentation may not reflect the actual work environment. Workshop attendees also expressed concerns about assumptions regarding future tools that may be necessary to support LDEMs, and stated that an intelligent system or personal assistant might be helpful in lieu of real-time support from MCC.

In a contextual inquiry research study to develop a list of high-risk, critical tasks for future deep space missions, Holden et al. (2020) identified spaceflight expert concerns that, “Without MCC, [the] crew does not have [the] knowledge or capabilities to address and solve every single vehicle issue that can arise.” (p. 6). “Response capabilities to unknown anomalies need to be enhanced; procedures need to be made customizable and adaptive. Intelligent systems that automatically give system status information following procedure completion would be a good idea.” (p. 5) and, “Because autonomous crew will not have consistent, timely access to MCC, autonomous decision-making capabilities need to be considered and enhanced.” (p. 7)

In addition to support from fellow experienced crewmembers, current ISS training also assumes that crewmembers will have ground support for more complex operations, such as for ultrasound imaging. According to a study by Kirkpatrick et al. (2013) on trans-Atlantic remote mentored ultrasound that included NASA medical operations remote guiders, ultrasound imaging is very “user-dependent, a characteristic that has prompted the development of remote guidance techniques, wherein remote experts guide distant users through the use of information technologies.” During ISS crew debriefs, crewmembers have commented that such remote guidance was needed despite having received preflight ground training on ultrasound techniques (FCI ISS crew comments database, n.d.); additionally, crewmembers have suggested that for LDEMs either a software solution or training solution should be considered, because the crew would not have a remote guider. Such training and performance support tools do not currently exist and will need to be developed.

Designing for Trainability

When designing tasks for trainability, NASA has the unique challenge that simulated environments and ground-based full-scale models or mockups cannot completely represent o spaceflight conditions. Representing a true μ-g environment on the ground has presented many challenges for training, so current simulations facilities and methods may not be adequate for preflight training. Figure 32 and Figure 33 depict training cardiopulmonary resuscitation (CPR) in a 1-g ground-based mockup and a demonstration of the trained operation being performed in μ-g onboard the ISS. Figure 33 was tweeted by ESA astronaut Samantha Cristoforetti tweeting that you, “Can’t train that on Earth” (Cristoforetti, 2014).

Figure 32. ISS Crewmember Satoshi Furukawa, Mike Fossum, and Sergey Volkov Train on CPR in a 1-G Ground Training Facility, Pushing Down onto a Dummy Strapped onto the Crew Medical Restraint System (CMRS)

Figure 33. ISS Crewmember Oleg Novitskiy Demonstrates One Method of Performing CPR in 0-G On Board the ISS by Pushing off the Ceiling with His feet and Down onto the CMRS

More critically, failing to design for trainability can lead to tragic consequences. In the NTSB report that investigated the fatal SS2 accident, the NTSB noted that,

> "Humans are susceptible to making mistakes regardless of the type of vehicle being operated, and system designs that account for the possibility of human error can help ensure safer vehicle operations. However, Scaled did not consult with a human factors expert (who would have the training and education to understand the interactions

> between humans and systems) to determine whether SS2's design would minimize the possibility of human error. Several Scaled engineers and a Scaled test pilot stated that they had taken a college-level course or had professional experience in human factors, but Scaled did not have a human factors department or a dedicated human factors expert on the company's staff. Scaled's vice president/general manager stated that the company addressed human factors by relying on input from pilots to identify and resolve ergonomics and other human factors issues, and Scaled's former chief aerodynamicist stated that the company relied heavily on input from its pilots because they and the engineers "made a vehicle that worked really well." (NTSB, 2014a, p. 46).

The NTSB report clearly implied that the Scaled engineers did not follow best HSIA practices in designing for trainability, including the fact that Scaled did not have a "dedicated human factors expert" on staff. Without this level of expertise in understanding human capabilities and limitations, the engineers designed a vehicle that allowed a single human command to result in a catastrophic failure, and, as described in Training Contributing Factor 1, the program of training and level of simulator fidelity was not sufficient to compensate for this design limitation.

Summary of Spaceflight Evidence of Training Contributing Factors to the HSIA Risk

The spaceflight evidence presented here clearly indicates that without an effective program of training that supports the retention and transfer necessary for in-mission performance and the level of expertise needed to operate a complex vehicle and respond to unanticipated critical malfunctions, there is a likelihood of adverse performance outcomes including that crew are unable to adequately respond to unanticipated critical malfunctions or detect safety critical procedural errors. To provide future crewmember with the training needed for autonomous and semi-autonomous mission operations research is needed to

- ➢ determine the retention interval for the skills trained to crewmembers of deep space missions, especially for mission critical tasks.
- ➢ develop guidelines and standards for the level of simulators and simulations fidelity necessary for skill acquisition, retention, and transfer, including simulator training for expertise with automation (such as robotics, intelligent systems, and decision-making tools).
- ➢ develop a systematic method for allocating duties and tasks from a learning perspective, while also considering teaming and behavioral health implications of such duty and task allocations.
- ➢ design a program of training across the entire training continuum that supports the level of retention and expertise needed for autonomous mission operations.
- ➢ determine methods that support both retention and transfer, and guidelines to ensure these methodologies are effectively employed.

- define evaluation measures for future missions that ensure not just skill acquisition, but retention and transfer.
- develop guidelines for performance support tools for in-mission task performance, including decision-making in response to unanticipated critical malfunctions.
- develop HSI design principles to support trainability of in-mission tasks.

Risk of Inadequate HSIA – and Mission, Process and Task Design Factors

Requirements, Policies, and Design Processes

Workload

An example of workload as a risk factor comes from the collision in June 1997 between the Russian spacecraft Progress 234 and the Mir Space Station, which caused the pressure hull to rupture and nearly led to the Mir being abandoned. Heavy workload and stress due to repeated system failures likely contributed to reduced vigilance capabilities (Ellis, 2000).

Numerous examples of heavy workload are evident in the Crew Notes database. For example, challenging physical workloads re imposed on the crew when they are expected to simultaneously wear protective gear and handle hazardous material while following the steps of a procedure or rotating a rack, or when coordinating with ground support and dealing with umbilicals. The Crew Notes database documents challenging mental workloads imposed by attempts to decipher unclear or convoluted procedures, when the nomenclature used on a procedure is not consistent with terms used on ISS labels or contain too many acronyms.

Operational Tempo

Stowage is one example demonstrating where 1g training provides an inadequate representation of information needed for task execution in 0g, and the time required to locate and access stowage items significantly impacts operational scheduling and tempo. As documented in the FCI ISS crew comments database, a true representation of the stowage of equipment and materials on board the ISS is very difficult to achieve on the ground and can create issues for the crew. Stowage mockups in 1g are limited because gravity restricts operations, translation, and stowage placement in the training facilities. Given the constraints of a 1g-based translation path, it is not possible or safe to place stowage items where they would potentially be stowed on board the ISS. On-orbit, the benefit of weightlessness allows stowage of items on any axis with proper restraint. The crew can translate through the available volume and position their bodies to move around obstructions or protrusions in the translation paths. Additionally, some of the stowage lockers are tightly packed, making it difficult to re-stow items on-orbit, due to the lack of gravity working against the crewmember. Similarly, crewmembers often have trouble with items floating off during retrieval or re-stowage. In a 1g environment, stowage does not behave the same as in a 0g environment, however, tasks and procedures may be written based on what is known from testing in a 1g environment and training in a mock-up. Given the gravitational differences between Earth and orbit, as well as the operational disconnect between ground training and life on-orbit, the crewmembers often experience steep learning curves once on board the ISS. Stowage guidelines have yet to be developed for LDEMs; new tools and stowage management methods must be researched for use during these missions.

The Crew Notes also provide operational evidence for the need to improve operations tempo on board ISS. For example, in developing a task timeline, schedules may not account for the crewmember's inexperience with the particular task or procedure or may not allow adequate time (even for seasoned crewmembers) to read the execution notes, scan the procedure to understand the overall goal of the task, collect required items, and set up the apparatus. This constant state of needing to catch up to the schedule can result in inadvertent errors.

Crewmember's failure to recall procedural knowledge they were trained on is another issue impacting operational tempo. Failures to recall increase with the occurrence of extreme circumstances, such as off-nominal conditions, stress, and heavy workload (e.g. Progress 234 and the Mir space station).

Procedural guidance

According to the FCI ISS crew comments database, crewmembers have been using paper checklists to assist them with their required on-orbit tasks for many years. However, navigating through paper checklists has been difficult. Procedures are coded with specialized symbols, abbreviations, boundary delimiters, and spatial configurations that collectively require extensive training to decipher and understand. Individual instructions in these checklists frequently take the form of conditional expressions (IF-THEN-ELSE statements), which the crewmembers must evaluate by manually crosschecking systems or flight status information on cockpit instruments and displays. The outcome of the evaluation of the logical expression determines which path should be taken through the remainder of the checklist, and that path, in turn, determines which subsequent instructions must be carried out and in what order. Wrong choices based on inappropriate assessment of the information presented can cause a risk to crew safety (Hudy & Woolford, 2008).

Crewmembers must use an extensive amount of information daily (Figure 34). Some ISS electronic procedures and formats have been especially difficult to use. The Crew Notes provide evidence that crewmembers frequently spend excessive amounts of time navigating between various menus because the procedures are difficult and lengthy, or occasionally because they contain unnecessary information. As documented in the FCI ISS crew comments database and in the Journals study, the structure and content of procedures have contributed to inadvertently skipping steps in the procedure that then resulted in poor task execution. In general, both the Crew Comments and Crew Notes databases contain entries suggesting that procedures are thought to be too detailed, especially for simple operations. Pictures and diagrams, considered helpful for many procedures, are not always integrated appropriately. In addition, the crew comments database documents cases where some procedures reference multiple steps from other procedures. Locating the necessary steps costs the crew additional time and has resulted in missed or skipped steps. Overall, usability of procedures has been an ongoing issue for the ISS and emphasizes the need for common human factors standards and simplification where possible in procedure development (Baggerman et al., 2004).

Figure 34. Cosmonaut Fyodor N. Yurchikhin, Expedition 15 Commander Holds a Camera While Looking Over Procedure Checklists in the Zvezda Service Module of the ISS

Crew performance of tasks on the ISS, such as EVA, maintenance, and medical tasks, relies heavily on the provision of adequate procedures and ultimately a strongly defined operational concept (Rando et al., 2005). The Crew Notes database provides evidence that procedures are often missing a step, have steps in the wrong order, graphics in the wrong location or missing information in a step (such as a missing specification). The astronaut Journals study also describe poorly vetted procedures. During ISS debriefs, crewmembers have reported that poor design of ISS procedures has impeded task performance by preventing the completion of scheduled activities within the allotted time. Well-designed procedures play a critical role in ensuring optimal on-schedule crew task performance; inadequately structured procedures will lead to a reduction in human task performance.

7 GROUND-BASED EVIDENCE

Much of the ground-based evidence for the risk of inadequate HSIA comes from aviation research and accident reports from other domains because the number of commercial, military, and private flights each year far exceeds the number of spaceflights. In addition, some of the evidence consist of summaries of subjective experience data, as well as non-experimental observations or comparative, correlational, and case studies.

Risk of Inadequate HSIA and Vehicle / Habitat Design Factors

Ground-based evidence includes opinion and observational and experimental data collected in settings such as laboratories and spaceflight analog environments. The relevance of this evidence varies depending on the environment in which it was collected, but all the evidence provided here is deemed to be useful in partially addressing habitability-related risks. Additional research will be needed to address the needs of future LDEMs.

Anthropometric and Biomechanical

Repetitive Stress Injuries (RSI)
Repetitive stress or strains over extended periods of time can affect the quality and quantity of productivity. RSI is a broad categorization encompassing a wide variety of injuries that are not always immediately apparent. The design of tools and workspace can cause some of these types of injuries. Sanders and McCormick (1993) describe human factors issues relating to cumulative traumas and repetitive strain. Although strains or injuries may not be mentioned in accident or injury reports, they can contribute to reduced work output, poorer-quality work, increased absenteeism, and single-incident traumatic injuries. Human factors reduce RSI through tool redesign and training of proper tool use. The authors emphasize that the tools should not be designed without considering the evaluation and design of the proper workspace, workstation, and task flow. During spaceflight, tasks such as glove-box operations, robotic operations, and some maintenance activities may require awkward postures that may impact tool design. Information is required to determine if any documented cases of RSI-related space injuries are attributable to the physical space environment. Repeated exposure to poor postures, constrained postures, and difficult and sustained moderate exertion can also lead to irreversible cumulative trauma disorders. Some of these persisting, poor physical ergonomic conditions could further be exacerbated during spaceflight. Knowledge is lacking on the effects of wearing a pressure suit while performing simple to complex tasks under pressurized conditions, possibly in a de-conditioned state due to exposure to microgravity.

A 2004 study of EVA training injuries and related symptoms in the Extra-vehicular Mobility Unit (EMU) suit found that hands, shoulders, and feet sustained the most injuries (Strauss, 2004). Most injuries and symptoms involved the hands (41% of reported cases), with the primary complaint being fingernail delamination (onycholysis) caused by axial loading of the finger coupled with excess moisture in the glove. The axial loading is attributed to extended reaching and forceful grasping motions during training. Glove fit and the design of the glove itself caused these injuries. Complaints regarding the shoulder (21% of reported cases) resulted from hard contact between the shoulder and the hard upper torso (HUT) of the suit, most often occurring when the crewmember is in a head down (inverted) position with the arms abducted from the body. Many crewmembers have experienced shoulder rotator cuff stress and strain injuries as a result. The Shoulder Injury Tiger Team determined that the EMU Planar HUT shoulder joint increases the risk of shoulder injuries due to the placement of its bearing relative to the body's shoulder joint (Williams & Johnson, 2003). The fit of the individual within the HUT and the design of the suit itself caused the symptoms and injuries. The feet were also affected (11% of reported cases); specifically, uncomfortable compression of the top of the foot and impingement of the distal toes attributed to problems with boot fit. The remaining symptoms concerned the arms, legs, neck and trunk areas, and all involved abrasions and contusions from contact with suit components. Strauss (2004) recommended increasing strength training for the hands and shoulders to minimize the potential for injury, and optimizing or improving suit fit for all reported areas of complaint. In summary, strength and anthropometry are critical components in the comfort and performance within the suit.

Spinal Elongation

Spinal elongation has a significant impact on crew performance because it can negatively affect crew and hardware fit, especially suit fit. Increases in torso length due to spinal elongation may cause suit discomfort or injury, or alter placement of hardware such as display location with respect to eye point.

Ground-based bed rest studies have also documented spinal lengthening. One bed rest study examined 8 male subjects after 16 days of 6 degree head-down tilt and found stature changes of 2.1 ± 0.5 cm (Hutchinson et al., 1995). Researchers noted that the spine lengthening observed during bed rest was not as great as lengthening during spaceflight because gravity was still acting on the spine of the bed rest subjects and they experienced less fluid shifts than astronauts in microgravity. An additional bed rest study used 6 degree head-down tilt as well as horizontal bed configurations for 6 subjects, who experienced 1.2 ± 0.1 cm increases in stature after 1 day of bed rest, and 2.2 ± 0.8 cm after 3 days (Styf et al., 1997), and length from L5 to C7 increased 2.5 ± 0.8 cm after 3 days. Researchers concluded that spine elongation is primarily responsible for changes in stature, further indicating that 30-60% of the change is due to increases in inter-vertebral disc height, and the remainder of growth is due to decreases in the curvature of the spine.

Neutral Body Posture (NBP)

Multiple experiments have been performed underwater to examine if a stable, replicable, and constant body posture occurs during relaxation (Dirlich, 2010). The subjects in one experiment performed up to 10 sets of repeated experiment trials (effort phases). Each effort phase started with one of 3 standard physical (muscular) effort tasks (e.g., full body crouch) selected by the experimenter without the subject's knowledge. After the performance of this task, the subject relaxed his/her muscles, and most subjects do relax to a stable, replicable, and nearly constant body posture during the relaxation phase. Although a large range of variability of postures existed between subjects, for almost all subjects, distinct goal-oriented movements were observed during relaxation, converging to a neutral resting position.

Other research on NBP in the Neutral Buoyancy Laboratory (NBL) indicates that variations found among subjects were significantly different than the results found in the Skylab study (Brown, 1976). The researchers found that NBP may actually change depending on the task a person is doing, and hence the task must be considered when considering NBP for design; however, this study used water to simulate microgravity. No studies have been conducted to date to understand the impact of NBP on suited operations.

Hardware and Task Considerations for Anthropometry and Strength

Sharp declines in strength have been reported for gloved performance when compared with ungloved strength. On average, subjects wearing an unpressurized Phase VI glove with a thermal micrometeoroid garment could produce only 55% of their bare-handed force, and 46% in the pressurized condition (Mesloh et al., 2010, July 17 - 20). Without the thermal micrometeoroid garment, gloved grip strength increased to 66% of bare-hand strength in the unpressurized condition and 58% in the pressurized condition. These effects result in reduced dexterity, tactility, and mobility during EVA missions, potentially increasing the duration, fatigue, and chance of injury to an astronaut. The EMU suit also reduces strength (Gonzalez et al. 2002), which may affect crewmembers' capabilities during certain flight scenarios.

The suit also restricts the range of motion. England et al., (2008) assessed how the suit affects tasks the astronauts were likely to perform during a lunar EVA, including walking, crawling, manipulating cargo, rotating a hatch, climbing a ladder, ingressing a recumbent seat, and more. The intent was to determine design requirements based on the minimum range of mobility necessary to perform realistic tasks, rather than imposing requirements for a full range of mobility while suited. The results have driven current HSI requirements to more accurately capture the reduced mobility of the suit, allowing designers to account for suited mobility (NASA, 2009a).

The EMU spacesuit causes performance decrements in comparison to unsuited operations. Capturing the effects of the suit allows these decrements to be quantified to ensure that all suited operations can be performed by all crewmember without risk of injury or risk to mission success. Unfortunately, the effects of the EMU on strength and range of motion have not been fully explored and based on experience these are highly dependent on the overall fit of the individual within the EMU, as well as the type of EMU the crewmember is wearing.

Visual Environments

Many of design aspects of lighting on a space vehicle are based on an understanding of the proper design of lighting systems for ground-based systems. The Illumination Engineering Society provides recommendations regarding lighting considerations such as illumination levels, color temperature, flicker, and system design (DiLaura et al., 2011).

Color accuracy is very important for color critical tasks (e.g., litmus test) and human perception of the health of their environment. If the spectral output of a light source does not render colors the users expect, there may be problems if they need to match colors. Additionally, color matching tests designed to work under one spectral light source may not work correctly under a different spectrum. Color accuracy may also be affected by a lack of uniformity of chromaticity and color temperature.

Flicker created by light sources that modulate their light output either by design or physical limitation can be distracting and potentially harmful to humans (U.S. Department of Energy, 2013). Flicker can create ghosting effects and can also trigger migraines in some people. The shape of the flicker modulation is just as important to consider as the frequency of the flicker. Lighting should be designed to eliminate flicker to minimize impacts to human health and performance. An architectural lighting system, including dedicated task lighting, is poorly designed if it does not consider how the occupant plans to use the space, the materials that are part of the space, and the room configuration. A “one light fits all” solution may not support a variety of uses and work space geometries. Designers who implement lighting without consideration to ambient and dedicated task lighting levels, spectrum requirements, variability of controls, flicker, interface with displays/monitors, windows (sources of usable light but also glare), and power constraints, run the risk of developing an incompatible design.

Vibration and G-Forces

Very limited research has been conducted to determine the effects of vibration and acceleration on human performance because studies typically require highly specialized, expensive test facilities. Deficits in vision, speech, and manual performance under vibration, and the health risks from acute and prolonged vibration exposure, have been widely documented for the 1-G loading of the Earth environment (See Griffin, 1990 for an extensive review of the literature). The preponderance of data pertaining to human performance under vibration were obtained from individuals primarily in upright and seated postures (Griffin, 1990). Operationally, some issues have been document with aircraft vibration during evasive maneuvers or system failures. During a British Midland Airways flight in 1991, the aircraft suffered a significant degradation of performance and propeller icing, accompanied by severe vibration that rendered the electronic flight instruments partially unreadable. The aircraft stalled, but control was regained and the flight ended safely (Air Accidents Investigation Branch [AAIB], 1992). In the laboratory, vibration studies involving semi-supine subjects, some dating back more than 50 years, have been conducted principally by or for NASA. These early studies informed the design of the spacecraft and the helmets for the Apollo and Gemini missions. Follow on studies have aided in the development of modern spacecraft such as the Orion capsule. Most of these studies were conducted with vibration superimposed on the Earth's 1-G bias, and a few were conducted on centrifuges that replicate the elevated G-loading experienced during space launch. In general, early studies found that as the vibration amplitude increases performance worsens, and the degree of degradation depends on the vibration frequency, direction, visual task resolution, and the seat-to-occupant interface.

During the Constellation Program (CxP), the erstwhile planned successor to the Shuttle Program, the issue of excessive, narrowly tuned *x*-axis vibration once again drew concern, this time because of the solid rocket motor thrust oscillation predicted for the first stage of the Ares-1 launch vehicle. Although the 10- to 13-Hz thrust oscillation band for Ares-1 overlapped the 11-Hz frequency that was studied intensively for the Gemini Program, crew interface concepts had since evolved considerably from the "steam-gauge" dials and incandescent lamps of the 1960s. Orion crewmembers will access information from a flat-panel, programmable "glass" display, akin to those of modem tablet or laptop computers, with capabilities for reconfiguring high-density text and presentation of graphical information. These advances in display technology necessitated revisiting issues of operator performance in a series of studies between 2008 and 2013 conducted at a 1.0 G_x bias load in the Human Vibration Laboratory and at a constant 3.8 G_x on the 20-G Centrifuge, both at the NASA ARC. Like the 1960s studies, this more recent work enabled NASA to develop crew performance-based program requirements to limit the magnitude of narrowly tuned vehicle vibration to 0.7 g_x for an instantaneous peak and a more stringent root-mean-square 0.21 g_y when averaged over a 5-s window. These recent investigations have led to new CMs against decrements in vibration-induced visual performance, and have revealed physiological effects and health risks due to the interaction of vibration and G-loading.

Studies of the general population indicated that both number-reading and speech comprehension were impaired under conditions of *x*-axis whole-body vibration (Adelstein et al., 2013; Begault, 2011). Studies of astronauts, both in the lab and on the centrifuge, extended the general population studies and indicated that readability of a graphical display was similarly noticeably disrupted under vibration. These and other recent studies examining newer graphical displays led the CxP to establish requirements that limited thrust oscillation vibration to a level that is slightly more lenient than the limit for Gemini and Apollo pogo vibration.

Because the Orion crew atop the current (i.e., Artemis) Space Launch System will experience the same 10- to 13-Hz vibration but this time in *y*-axis from trust oscillation of side-mounted solid rocket boosters, number-reading was studies with side-to-side whole-body vibration. In contrast with *x*-axis vibration, no objective effect of y-axis vibration was noted for number reading although subjective ratings of difficulty and workload grew with increasing vibration magnitude. As a result of this investigation, the legacy CxP requirement was revised for Orion to now also limit lateral thrust oscillation-driven vibration to an instantaneous peak of 0.7 g_y and the equivalent root-mean-square average of 0.5 g_y over any 5-s window.

As an alternative to what would have been costly vehicle modifications to reduce Ares-I's trust oscillation-induced vibration had CxP continued, Adelstein and colleagues (2013) at ARC devised and tested a display-strobing countermeasure that instead mitigated the visual blur and the consequent performance decrements caused by the vibration. Flashing a display panel's backlight in time with the seat vibration restored error rates for the number-reading task at 0.7 g_x vibration to the same levels as when the observer was not vibrating at all.

A series of lab and centrifuge studies of visual-manual performance at ARC demonstrated physiological and health interactions between x-axis vibration at 8, 12, and 16 Hz, and G-bias loads of 1.0 and 3.8 G_x. From June 2008 through August 2010, NASA researchers investigated whole-body vibration impacts on human cognitive and manual performance (Adelstein et al., 2009a, b, c; McCann et al., 2009; Sándor et al., 2010c). All subjects were unsuited and in a semi-supine position, representative of the body posture during a stack-architecture launch, and received single-axis vibration in the chest-to-spine direction (the expected dominant component of Ares-Orion thrust oscillation inputs). Most of these studies focused on the ability to read text on a display under vibration, although one focused on the use of cursor control devices (CCDs) under vibration. Objective measures indicated that text readability is not significantly impaired at vibration levels up to 0.3-g (0-peak) for 10- and 14-point Lucida Console font viewed at distances of 18-20 inches, and that display usability and workload are not significantly impacted. Vibration at higher levels may significantly impact visual performance; more research remains to be done.

Observations from centrifuge and fixed-base vibration chairs indicate that visual performance is degraded more by vibration superimposed on a 3.8-g chest-to-spine bias than a 1-g bias. Increasing font size from 10 to 14 points or larger may further improve text readability or permit acceptance of higher seat vibration levels. Incorporating appropriately designed, vibration-resistant graphic formats, as opposed to small-font dense displays, may also improve readability. Additionally, CMs such as appropriately designed display strobing may allow greater vibration tolerance for both text and graphics formats.

A study performed in the ARC Vibration Facility determined the characteristics of cursor control devices that perform well or are problematic under different levels of vibration (Sándor et al., 2010c). The platform provided one axis of vibration (x-axis/chest to spine) at various amplitudes and frequencies. Displays for the CCD tasks were shown on a monitor mounted in a fixed position at viewing distance above the participant’s head; the monitor did not vibrate. CCDs were mounted on the chair for left-handed use as planned for Orion (Figure 35).

Figure 35. Cursor Control Device Being Tested in the Vibration Chair

A trackball was tested in continuous and 4-way discrete mode, a castle switch in 2- and 4-way discrete mode, and a rocker switch in 2-way discrete mode. The higher amplitudes vibration (3Hz 0.17g, 6Hz 0.35g and 12Hz 0.70g) affected performance of all devices. Lower amplitude vibrations did not cause a significant decrement in device performance. Response times for the 2-way devices were negatively impacted by vibration, whereas 4-way and continuous modes were not. Further research is warranted to confirm/clarify these results.

In 2017, NASA's Health and Medical Technical Authority for the Orion program exposed astronaut crew subjects to flight-representative, time-varying *x*-axis vibration. These investigations were unique in that they employed the very-high-fidelity Orion seat and display panel mock-ups and marked the first suited and helmeted (with visors down) assessments of display usability and coupled seat-occupant dynamics since the 1960s.

Recent non-programmatic research on visual-manual performance extended 1.0- and 3.8-Gx 12-Hz g_x vibration experience to *x*-axis vibration at 8- and 16-Hz; however, studies of performance during a specific single axis vibration scenario are limited in scope and are valid only for the specific design points examined. Performance needs to be assessed under more general broadband (i.e., multi-frequency) and during concurrent multi-axis vibration.

The performance impacts under G_x-loading (i.e., semi-supine) conditions of vibration frequencies other than those described above are generally not known. Although vibration and G-loading individually have been shown to raise heart rate, Godinez et al. (2014) reported that of the 3 vibration frequencies they studied, heart rates were highest for the 8-Hz vibration under $1G_x$ bias, and lowest at 3.8 G_x, a G-by-vibration interaction they ascribe to G-induced biodynamic changes that were originally noted by Vykukal (1968) and Vogt et al. (1973). Most notably, 3 participants over the age of 50 developed benign paroxysmal positional vertigo after centrifugation, whereas no cases were seen after vibration at 1.0 G in the lab (Liston et al., 2014). These clinical observations indicate a heightened risk of vestibular injury due to G-loading that increases with age. Because no incidents of benign paroxysmal positional vertigo were reported in the prior 12-Hz studies, the new vibration frequencies in these studies may have contributed. Systematic study of concurrent vibration and G-loading spanning a broad range of load combinations is needed to develop models predictive of human performance and health impacts and countermeasure strategies. Otherwise, analyses traceable to empirical data (i.e., validated models) will need to be developed on a case-by-case basis if they are to be used to assess cost-to--benefit trades between vehicle designs (i.e., vibration source/transmission mitigations) and human factors (i.e., displays, controls, seating and suit) design modifications. Without new research to fill these analysis, design, and verification gaps, vehicle programs will have to accept additional levels of risk or provide sufficient (and potentially excessive) margins to ensure occupant health and effective performance during launch and landing.

Noise Interference

Exposure to noise can negatively affect human health through temporary or permanent hearing, or can affect human performance through interference with task execution (Tayyari & Smith, 1997). Annoyance factors related to noise may also affect performance because they induce stress. As such, noise has been heavily studied in ground occupational settings such as manufacturing environments and commercial and military flight applications. The Occupational Safety and Health Association regulates workers' exposure to cumulative noise levels over a given time period and to impulsive noises (Occupational Safety and Health Administration). The American Conference of Governmental Industrial Hygienists also determines threshold limit values for noise pressure levels (American Conference of Governmental Industrial Hygienists 1993). These overarching understandings of noise in the environment must be considered for spaceflight, and the use of modeling and HITL evaluations are pivotal in ensuring an adequate spaceflight environment with respect to noise.

Seating, Restraints, and Personal Equipment

Seats must be designed to accommodate the dimensions of the crewmembers, and must be adjustable to allow for body movements, reach, access, and position during the dynamic phases of flight (e.g., launch and landing) (NASA, 2011). It is challenging to provide sufficient occupant protection without severely compromising the crews' ability to reach controls and turn their head to see displays. Seat design is also often complicated by the need to accommodate suit appliances such as umbilicals, or seat-mounted devices (hand/cursor controllers), as determined during HITL assessments in Orion mockups.

Crewmembers wear a flight suit and are seated in a recumbent position during launch and entry. Wearing a flight suit while restrained in a seat may significantly affect anthropometry, and recumbent seated anthropometry can differ from upright seated anthropometry. These considerations are critical for optimizing space vehicle design, and when considered early, ensure lower design retro-fit or redesign costs (NASA, 2011). When seating is designed correctly, the vehicle's cab accommodates a diverse crew, performance is optimized, and safety is increased.

There is a risk of excessive crew discomfort during long-duration seated periods due to the seated posture and lack of lumbar spine support in vehicle seats. During some early CxP-led HITL evaluations of the Orion vehicle, several test subjects complained of lower back discomfort and numbing of the feet and legs while seated in the mockup seats for long-durations. This could be a problem for most crewmembers during pre-launch activities. Paresthesia resulting from the long-duration static seated posture could inhibit emergency egress, endangering the crew. Also, excessive discomfort can cause an increase in workload, which could result in an increase in task error rates and a decrease in task performance.

The Orion Crew Impact Attenuation System evaluation indicated that the seat design requirements had been met if the commander and pilot seats were configured for crewmembers with short sitting heights and medium to large buttock-to-popliteal dimensions, however, insufficient clearance existed for ingress or egress of the seat because the seat pan was too close to other cockpit vehicle hardware. This example highlights 2 important risks. First, it shows that individually designed components can meet requirements when tested separately from the system but fail to meet the final design intent when integrated with the human and hardware system. Second, this example illustrates issues with the way human accommodation and population variation is typically handled during the design process. Recently, digital humans have become a standard tool to quickly and inexpensively assess how a design will accommodate users; however, assessments typically only include minimum and maximum manikins, representing the population of interest, in the virtual space so that reaches, clearance, and interactions can be calculated. This min/max modeling method was used for clearance calculations of the Orion ingress/egress space, but because the clearance issues only affected specific portions of the population, they were not identified in the analysis. If digital human modeling is not coupled with HITL evaluations, some of the integrated system design issues may not be discovered until late in development when changes are very costly. In this example, early Orion HITL evaluations were clearly valuable in addition to model-based analysis.

Visibility/ Window Design & Placement

Windows are important in design of long-duration habitats on Earth, and astronauts often mention the positive aspect of windows (Stuster, 1996). Both health and psychological reasons exist for providing windows in a vehicle or habitat. The absence of distant objects on which to focus is thought to contribute to submariners developing temporary esophoria, and their distance vision may permanent degrade over the course of their career (Kinney et al., 1979). In addition, the ability to “periscope liberty”, enabling glimpses outside of the confined living space, significantly increases submariners’ morale (Weybrew, 1979).

Vehicle/Habitat Volume/ Layout

Insight into spaceflight needs for habitable volume and layout may often be gained through ground-based studies, and from the experience of industries that deal with confinement and isolation. Recent relevant ground-based studies include work at the NBL, and a Russian-led isolation study. NASA has coordinated multiple collaborates with external experts from industries such as oil and gas, mining, maritime shipping, and submersibles regarding their approaches to designing for habitable volume needs. Examples of results from these efforts are described below.

Analog Studies

Exterior design of the spacecraft affects the translation, efficiency and safety of external EVA spaceflight activities. Hatch opening, ingressing the hatch, handrail removal, closing the hatch, and handrail translation tasks were examined in a joint CxP/ISS integrated test conducted with participation from the EVA Systems Project Office and the Orion Project Office (DeSantis et al., 2011). The test was conducted in the NBL with subjects wearing EMU suits. Ratings were provided using a usability ratings scale, Cooper-Harper Scale (for overall handling qualities of the suit), and the NASA-developed Maneuverability Assessment Scale for 3 scenarios: (1) ISS-based EMU EVA (i.e., translation from the ISS node handrail to the Orion side hatch); (2) Altair-to-Orion Transfer (i.e., translation from the Altair vehicle handrail to the Orion side hatch), and (3) Orion-based EVA (i.e., translation from inside the Orion side hatch to remove EVA handrails). Locations for needed handrails were identified, and for some hatch activities, the lack of a defined handhold affected participants' ability to close a hatch. For example, to close the hatch, all subjects in the ISS-based EMU EVA had to egress the hatch almost out to their waist to grab onto an indentation, and one subject (with the shortest arm span) was unable to complete translation, hatch opening, or hatch closure tasks. The limitations of the EMU suit architecture could have adversely affected smaller participants. Hatch ingress and hatch closure activities involved inadvertent contact with hatch seals, which can negatively impact the seal's effectiveness and thus pose a life-threatening risk to the crew. Habitat components (e.g. hatches, handrails, and translation paths) must be designed to consider both the human limitations (e.g. suited small arm span) and components that could be affected by potentially repeated contact (e.g. hatch seals).

A recent 520-day isolation study simulating a mission to Mars highlights the relationship between the environment and behavioral outcomes. Basner et al. (2013) found that 4 of the 6 crewmembers experienced one or more disturbances of sleep quality, vigilance deficits, or altered sleep-wake periodicity and timing, suggesting inadequate circadian entrainment. Crewmember sedentariness also increased across the mission. Such behavioral changes indicate that ecological variables can be determinants of sleep. Lack of exposure to the Earth-day light cycle, and the lack of light enriched in the "blue" end of the spectrum, the most potent region for maintaining circadian rhythms and promoting sleep, may have led to such changes. These findings demonstrate the critical role of the environment in maintaining behavioral health and performance.

Net Habitable Volume (NHV) Workshops

Although some studies suggest that volume and layout of a spacecraft may mitigate some psychological stressors during long-duration spaceflight, the data are limited. Experts from NASA, military, and academia discussed this topic at the 2011 Habitable Volume Workshop, and agreed that the lack of tests beyond 6 months of confinement limits the amount of applicable information required to determine NHV (Simon et al., 2011). The following areas of forward work were identified: prioritizing the knowledge related to NHV needs and documenting steps for mitigating gaps; developing methods or test beds allowing for future testing on ISS; identifying practical metrics and tools for assessing layout; and conducting a long-duration confinement and isolation study of minimum acceptable NHV for LDEM durations.

During a follow-on workshop in 2012, experts sought to create products to help determine the minimum acceptable NHV (Thaxton et al., 2014). Participants in the 2012 Habitable Volume Workshop included NASA stakeholders, and academic and industry experts from fields such as oil and gas, Navy, maritime shipping, submersibles, and mining. Products of the workshop included a high-level process to design for minimum acceptable NHV, a list of volume-driving tasks for LDEMs, and a preliminary list of metrics and methods useful for assessing habitability. Several areas of forward work were identified, including better definition of volumes needed for specific tasks; refinement of modeling and simulation approaches; definition of behavioral health and performance impacts; and collection of spaceflight data.

Based on discussions at the HRP-sponsored Habitable Volume Workshops (Simon et al., 2011; Thaxton & Chen, 2013), it is evident that the community of SMEs believes that habitable volume needs for long-duration missions differ from habitable volume needs for short-duration missions. Agreement was reached that further research is needed to quantify those differences. For example, differences in perceptions of privacy or changes in team dynamics may exhibit later in the mission. Alternatively, research may show that a 1-year mission does not differ widely from a 6-month mission in terms of effects on habitability and human factors. Thus, a clear need exists to perform this type of research during a ISS 1-year mission and build on this initial work to gain a better understanding of issues that can be examined in a microgravity environment.

In January 2014, the HRP hosted an NHV Consensus Session (Whitmire et al., 2015). This effort consisted of ‘grouping’ like tasks expected on a LDEM into functional areas, and then using previously defined task-based volume numbers to determine the volume needed for that multi-functional area. The key functions of a crew on a Mars mission were mapped to hypothetical, physical locations on a vehicle. The group of SMEs abided by mission assumptions and created a layout that provided 25 m^3 per crewmember (for a crew of 6 people), which includes 5.4 m^3 for individual quarters (as a protective factor relative to the extended duration of the mission and the small, confined space of the remainder of the vehicle). This estimate is similar to the “starting point” value of 26.85 m^3 per person provided in the Human Integration Design Handbook (NASA, 2014). Assumptions affecting the recommended NHV during this consensus session include no need for airlocks or suit ports; adequate acoustic isolation between disparate activities; adequate space in common areas to accommodate all crewmember simultaneously; adequate sensory stimulation; and ability for crew to optimize aesthetics and take personal control over aspects of the environment. The volume estimate was based primarily on defining functional zones for tasks and estimating the required volume for the tasks that are assumed to take place within that zone (e.g., dining and communal activities, exercise, and hygiene).

Risk of Inadequate HSIA and Human Computer Interaction (HCI) Factors

In this section, the 7 factors used in the HCI Spaceflight evidence section will be related to ground-based incidents. Although most of the evidence reported is from the aviation and ground transportation domains, inadequate HCI is a factor in many other realms, such as healthcare, multiplayer video games, emergency response, and augmented VR.

Informational Resources/Support

Numerous incidents exist from multiple domains of lack of information in the required format or lack of critical status information. In 1987, 188 people were killed in a ferryboat accident in Zeebrugge, Belgium. When leaving the dock, the bow doors had been accidentally left open, and the ferry filled with water and sank. In addition to poor assumptions about whose responsibility it was to ensure the doors were closed, the captain had no status indicator on his console indicating the status of the doors, and the doors were not visible from his location (Casey, 1993; Pijnenburg & Van Duin, 1991).

Lack of indicator was the root cause of another accident in 1985. A Galaxy Airlines aircraft crashed when ground crew failed to secure an air-start equipment door on the airliner wing prior to takeoff. During flight, the door came loose, but pilots were unaware of the problem because there was no indicator in the cockpit. The plane became unstable and ultimately crashed, killing 70 passengers (Chiles, 2002; National Transportation Safety Board [NTSB], 1986).

Indicators or safeguards were not in place to prevent the accident of Aeroflot Flight 693 (Accident Description - Aeroflot Russian International Airlines, Airbus A310-304, 2004). In 1994, the plane stalled and crashed, killing 75 people. The autopilot was accidentally disengaged by the pilot’s 15-year-old son sitting in the cockpit.

Inadequate presentation of information was the cause of another plane crash in 1983. The Korean Air Lines Flight 007 deviated more than 200 miles from its intended flight route into Soviet territory and was shot down with no survivors (Degani, 2001). The deviation from flight route was primarily related to the pilots’ use of automated systems. The autopilot provided inadequate information about its mode transition logic and did not provide adequate information to flight crew about its active and armed modes. The autopilot displays were incomplete in that they did not accurately represent the currently engaged mode. In this case, poor IA, information presentation, functional logic, and the pilot's inadequate understanding of the system interacted to create a disastrous condition.

Critical information often involves more than just visual displays. In 1999, during London's morning rush hour, a commuter train collided with a high-speed train, killing 31 people. An investigation revealed 2 HCI-related issues: (1) minor warnings and critical events both had the same type of auditory alarm, and (2) when the pilot acknowledged the alarms, the system cancelled them altogether, eliminating a layer of protection (Cullen, 2001; Lawton & Ward, 2005). In this case, there was lack of information regarding severity of the event (since alarms used the same tone), and lack of information regarding a continuing threat because the system allowed the user to essentially disable the warning system by acknowledging the warning.

Multiple studies (Begault et al., 2008; Begault et al., 2007) at NASA have evaluated auditory alarms. The studies showed that inclusion of a speech suffix along with the alarm tone was preferred by both crew and non-crew operators. Two later studies (Sándor et al., 2010a, Sándor et al., 2016) investigated how quickly auditory alarms with and without a speech suffix could be identified. The latter study also included an evaluation in the HERA, a habitat facility at the NASA JSC used for simulating spaceflight missions. On average, the participants responded faster to speech than just alarm tones. The response time to react to an alarm will be essential during LDEMs when MCC may not be available to assist the crew in real time to detect, diagnose, and mitigate an emergency. Although these studies provide preliminary data, clearly, additional studies are needed to develop standards and guidelines regarding how auditory alarms should be combined with speech for best results. Specific guidelines will be needed for future space programs to implement these alarms.

In one of the worst industrial accidents on record, the 1984 Union Carbide India Limited pesticide plant in Bhopal, India experienced a toxic gas release that killed an estimated 3800 people at the time of the incident, and thousands more later from gas-related diseases. A highly critical pressure gauge that should have warned operators of an impending problem was missing from the control room (Casey, 1993). This gauge was located somewhere else on the plant site and was required to be monitored manually. In addition, another critical panel had been removed from the control room, perhaps for maintenance. Many of the gauges in the control room were reported to be consistently broken, malfunctioning, off-scale, and unreliable. Training was poor, signs and procedures were written in English, when many of the operators only spoke Hindi. This is a clear example of lack of informational resources such as proper displays, procedures, and communication, leading to a catastrophe.

Communication was also an issue in 1995 in the crash of American Airlines Flight 965 to Cali, Colombia. The crew experienced several difficulties including loss of waypoints from the navigation computer, and inconsistent and unfamiliar labeling on charts. The air traffic controller noted that some of the pilots' requests did not make sense, but they did not know enough non-aviation English to convey this. Navigation and communication issues resulted in the aircraft hitting a mountain, killing 159 passengers and crew (Ladkin, 1996).

Miscommunication with air traffic control was a primary contributing factor to the crash of the Avianca Flight 52 (NTSB, 1991). The flight was delayed en route numerous times by weather and was dangerously low on fuel as it approached New York in January, 1990. Air traffic controllers were unaware of the low-fuel situation and the plane crashed in Long Island killing 73 passengers and crew.

Inadequate procedural information can result in crews not having the information they need to perform a task. In his keynote address at the 2010 Human Factors and Ergonomics Society meeting in San Francisco, Captain "Sully" Sullenberger, the U.S. Airways pilot that guided 155 passengers and crew to an emergency water landing in the Hudson River in 2009 highlighted the need for emergency procedures with just the right level of information. Captain Sullenberger said that in the critical moment, he wished he had procedures that could collapse to contain just the key 4 or 5 steps required for emergency landing (Sullenberger, 2010).

Procedures with the right level of information are a good first start, but long-duration space missions require procedures that support SA for the astronauts while maintaining their performance (and reducing dependence on memory). Holden et al. (2018b) investigated level of procedure automation by assessing the effect of increasing level of automation (LOA) on SA and other human-system performance metrics. Although the study showed that LOA had no significant impact on SA, the manual procedure produced the highest workload, and high automation conditions produced the lowest workload. Similar results were found in a study by Oman and Liu (2018) who investigated how the allocation of procedural step execution between the human operator and automation, or LOA, would affect task performance, SA, and mental workload during a robotics task. Thus, more work is needed to understand the appropriate LOA with respect to electronic procedures and checklists to maintain and improve SA during LDEMs.

Gawande (2009) advocates for the use of checklists in the medical domain, similar to the procedures that are currently in use on the ISS and that must continue to be developed for use during LDEMs. Medical procedures have developed over the past century to a point of such extreme complexity that it is impossible for a single person, or even a team of specialists, to properly carry out all the necessary steps for proper patient care from memory alone. The average Intensive Care Unit (ICU) patient undergoes nearly 180 procedures in a single day that are often performed under pressure. Catheter-related infections in ICU patients are common, expensive, and sometimes lethal. Pronovost et al. (2006) implemented the use of 5 procedures previously identified by the Center for Disease Control as having the greatest effect on reducing the rate of catheter-related infections. Following the implementation, infection rates in ICU patients decreased dramatically, saving both lives and money. Checklists and procedures have also been successfully used in construction, restaurants, investing, and, of course, aviation. However, checklists and procedures must be carefully designed. If they are vague, imprecise, too long, they are impractical and unlikely to be used.

Voice communication in the context of online multiplayer games is similar to crew resource management in some aspects. Halloran et al. (2004) found that online voice communication makes it difficult for the player to know who is speaking to them, and this impairs performance in team-based games. Inadequate information or feedback is detrimental to accomplishing the goal, and additional information, possibly multi-modal such as text identifying the speaker, should be available to the user. Moore et al. (2007) identified issues with text-based communication in gaming contexts in which long gaps in the conversation led to confusion and miscommunication. This is similar to communication delays that may occur in LDEMs. Additionally, while the user may be engaged in other activities that are not visible to their partner (e.g. opening their inventory), their avatar, or representation in the game world, appears to do nothing. Again, lack of timely and appropriate feedback inhibited task completion and contributed to confusion and miscommunication.

In emergency situations, responders must be provided with the appropriate amount and type of information to aid their efforts. Communication systems must provide information to people who are not collocated, serve different roles (e.g. first responder at the scene versus someone at the command center), use different devices (e.g. high-resolution computer screens versus mobile device), have differing experience levels, and require different types of information. Context awareness of these communication systems is therefore crucial to ensuring that the right information is conveyed to the right people in an appropriate manner to facilitate smooth response to the emergency. This will also help reduce the cognitive load of the user by not overwhelming them with unnecessary information (Flentge et al., 2008). These factors must be considered in the development of emergency response communication systems, and should be considered when developing procedures and other systems for use during spaceflight.

Attention and Cognition

In 1995, pilots aboard American Airlines Flight 965 mistakenly cleared their approach waypoint from their navigation computer. The pilots attempted to reenter the waypoints, but due to an inconsistency in the landing site designation, entered the wrong site designation. The autopilot turned the aircraft, but at that point the aircraft was in a valley parallel to the desired one and on a collision path with a mountain. When the ground proximity warning system (GPWS) annunciated, the pilots were unable to clear the mountain, approximately 12 seconds before the plane collided with the tree line, killing 159 people on board. In a report by *Aeronáutica Civil* (Ladkin, 1996), the cited probable causes included “poor attention of the crew to vertical navigation, proximity to terrain, and location of critical radio aids”. The flight crew did not try to terminate the descent, although the airplane deviated from the published approach course. Nor did they retract the speedbrakes while performing an escape maneuver after the GPWS annunciation. The investigation committee concluded that had the speedbrakes been retracted, it is likely that the plane would have cleared the mountain. Poor allocation of attention by the flight crew was partially due to the crew trying to make up time due to delays before departure. In their hurried state, they performed an inadequate review of critical information and their attention was diverted from flight instrumentation. In the critical time after the GPWS annunciated and approaching collision, the pilots may have been unable to redirect their attention to the situation and may have experienced cognitive tunneling that prevented them from realizing that the speedbrakes were still engaged.

In 1986, the passenger plane Aeroméxico Flight 498 collided mid-air with a private jet over Cerritos, California, killing all passengers and crew on both aircraft and an additional 15 people on the ground. The private plane had flown into a controlled area without clearance. At the time, the traffic controller was focusing on another plane that had also entered the controlled area without clearance. No system existed to warn the air traffic controllers about impending mid-air collisions. The NTSB (1987a) concluded that the pilots of the 2 aircraft could have visually made contact through their out-the-window views, but did not. Thus, not only had the air traffic controllers' attention been diverted from the position of the private jet, but the pilots' attention may have been diverted from their out-the-window views, which would have allowed them to see each other's planes and possibly avoid the collision.

In other incidents, inappropriate allocation of attention to faulty flight instruments or indicators has resulted in inadvertent activation or disengagement of critical systems. For example, the flight crew aboard Eastern Air Lines 401 became distracted while trying to fix a landing gear indicator light that did not illuminate, which led them to accidently switch autopilot modes. Without the pilot's attention to altitude indicators, the plane descended and crashed into the Florida Everglades (NTSB, 1973). In a similar situation, the flight crew aboard Adam Air Flight 574 inadvertently disengaged the autopilot while trying to fix a problem with the inertial reference system. The pilot became preoccupied with troubleshooting the system and did not attend to information about the plane's increasing speed. By the time the situation became apparent, the pilot could not recover the aircraft in time and the plane experienced structural failure (National Transportation Safety Committee [NTSC], 2008). In both the Eastern Air Lines 401 and Adam Air Flight 574 cases, the pilot and copilots' attentional resources were focused on salient, but faulty, indicators that distracted their attention from flight instrumentation they would otherwise nominally monitor. The changes in altitude and speed when autopilot was disengaged or put into an alternate mode were not noticeable enough to warrant the pilots to redirect their attention to these flight instruments.

Failure to properly attend to automation has also been cited as the probable cause for other incidents. In 2009, Turkish Airlines Flight 1951 crashed into a field near Amsterdam during landing. As the plane was approaching the runway, a faulty radio altimeter triggered the autothrottle to decrease engine thrust. The flight crew did not notice the reduced airspeed until it was too late. They were not monitoring airspeed or altitude, and instead were relying on autopilot to land the plane. A 2010 investigation by the Dutch Safety Board (DSB) found that the pilots did not disengage the auto throttle and thus failed to take over manual thrust to increase airspeed. The crew was not aware of the decreased air speed, and failed to respond to it and to the "impending onset of the stick shaker", although there were several indications. Cues included a box around airspeed display that turned amber and flashed for 10 seconds, aural and haptic feedback, and a red-dashed "barber pole" stall warning indication. If the pilots had noticed these cues, they would have had enough time to recover and perform corrective action. The pilots, however, may have been distracted by having to perform a landing checklist for attitude below 1000 feet, instead of attending to flight instruments. Furthermore, the pilot did not communicate to the co-pilot that his altimeter was malfunctioning. Thus, the pilots failed to allocate attention to numerous pieces of information that would have indicated to them an approaching stall, and failed to direct their attention to faulty instrumentation.

Cognitive tunneling is a real possibility when managing a difficult fault during spaceflight (McCann & McCandless, 2003). For example, during a Shuttle ascent, which lasts approximately 8.5 minutes, the crewmembers perform checks of various time-critical parameters and flight instruments, and they must act quickly to assess and react to a fault (Huemer et al., 2005).

During HITL evaluations of displays for a future crewed space vehicle, participant errors occurred due to inappropriate allocation of attention between electronic procedures and system displays (Ezer, 2011). The decision support system provided by the electronic procedure engine, which cues up displays and commands for crewmembers, may lead to over-reliance on procedures. Design concepts are being considered that will encourage crewmembers to allocate their attention equally between procedures and associated system displays.

Individuals perceive the passage of time differently depending on their level of expertise at a task (Block & Zakay, 1997) and their cognitive load (Thomas & Weaver, 1975). Additionally, exposure to microgravity can alter internal timekeeping mechanisms (Semjen et al., 1998). Temporal awareness is especially important when a task requires the coordination of multiple individuals. Space exploration requires time-sensitive coordination between individuals in space and on the ground. Therefore, a better understanding of the effects of spaceflight on temporal awareness is needed to design interfaces that can assist temporal coordination within and between space and ground crew.

Attention deficits occur during spaceflight (Kanas et al., 2001), and in spaceflight analog environments (Basner et al., 2013) and studies of elevated CO_2 levels (Fothergill et al., 1991). For example, reduced alertness has been reported in Antarctic analogs (Mullin, 1960; Palinkas & Suedfeld, 2008), and reduced response to rare events during confinement (Mecklinger et al., 1994). A more complete understanding of these deficits is required to create displays with built-in CMs for mitigating the performance decrements during LDEMs.

Embrey et al. (2006) describe several maritime incidents that have resulted from cognitive overload. In 2004, the ferryboat Catherine Legardeur ran aground due to a combination of factors resulting in cognitive overload. The master of the vessel was accustomed to navigating using visual references from the environment. When fog reduced visibility of the environmental cues, he struggled with the unfamiliar navigation instruments. Due to high cognitive workload, he lost control of the ferry. In another incident in 2005, the skipper of the Hannah Lee was busy making continuous adjustments to an autopilot system that was not integrated with the electronic chart system. Spray on the wheelhouse windows partially obscured visibility. The limited visibility and the high cognitive load of the continuous adjustment task resulted in the skipper being unable to maintain a proper visual lookout; ultimately the Hannah Lee collided with another vessel, the Spartia.

Just like cognitive overload, cognitive underload can affect performance and safety. Embrey et al. (2006) describe several maritime examples. In 1999, the Baltic Champ vessel was anchored and the master was not engaged in any activity. He was rested and on watch alone in the wheelhouse. He walked around and checked radars to try and keep alert; however, because of the low activity level, he did not notice that the vessel was drifting. The Baltic Champ ran aground before the master could take action to prevent it. In another maritime example from 2002, the master of the Stellanova, acting as the officer of the watch, was doing administrative tasks while traveling through a narrow channel. When the Stellanova suddenly sheered to port, the master was unprepared for the higher cognitive load task of collision avoidance. He was unsuccessful in assessing the situation in time to avoid collision with the Canadian Prospector.

Cognitive load and attention are closely related to the concept of "flow". In discussing the development of educational games, Kiili (2005) emphasized the importance of flow, which is defined as complete engagement and absorption in the task. Antecedents to flow include focused attention, clearly defined goals, feedback, and ease of use of the system. All of these factors that promote flow in the gaming domain are also important for completing most tasks crewmembers are faced with today and will be increasingly faced with during LDEMs. Kiili suggests that bad usability limits flow because the user must focus on extraneous information that is irrelevant to the goal itself. Similarly, if the spaceflight task is very complex, and/or there are usability issues with the system, the user may experience cognitive overload, which may inhibit the development of flow and detract from the experience. Kiili suggests that to minimize cognitive load and enhance learning in a gaming situation, developers should consider including haptic feedback. This kind of feedback will support immersion and will not occupy limited streams of verbal and auditory working memory. Haptic feedback during spaceflight may minimize cognitive load and enhance crew performance.

Motor Skills/Coordination and Timing

Fine motor skills change over a person's lifetime and are influenced by several factors. Age-related decreases in fine motor performance are well documented. For example, when comparing a group of healthy adults over the age of 60 to a younger group, Smith et al. (1999) found that older adults had slower performance on fine motor tasks. Task difficulty also affected performance time such that a more difficult fine motor task resulted in longer completion times, more so for the older group than for the younger group.

In addition to the normal process of aging, certain disorders may also affect fine motor skills. Individuals with Alzheimer's disease (AD) and Mild Cognitive Impairment (MCI) had slower, less smooth movements in a handwriting task than did controls, indicating that AD and MCI may affect fine motor skills (Yan et al., 2008). Congenital hypothyroidism (CH) has been associated with long-term diminishing of fine motor skills: young adults with CH performed worse on a Grooved Pegboard task than did their healthy siblings (Oerbeck et al., 2003).

These studies show that fine motor skills change naturally during a lifetime and as a result of disease or abnormalities. Given that exposure to microgravity leads to a variety of physiological and possibly cognitive changes, fine motor skills may also be affected. The effects of prolonged exposure to microgravity must be examined before LDEMs.

Design of Displays and Controls

Fitts and Jones (1947) interviewed 100 pilots to identify errors made during reading and interpreting instruments. They found that the 2 most common errors were reversal errors in which interpretation of instrument readings, such as heading, were reversed, and errors in interpreting instruments that had multiple rotations, such as altimeters. Several pilots reported misreading the altimeter by 1,000 ft. Other frequent errors involved lack of instrument legibility and substitution errors in which the wrong instrument was viewed. One pilot reported that he was unable to fly a new model of an airplane during a Japanese bomber attack because of the inconsistency between the displays and controls of the new and old models.

In 1998, Swissair Flight 111 crashed into the Atlantic Ocean near Nova Scotia, killing all on board. The Transportation Safety Board of Canada (TSB) concluded that emergency gauges were in awkward positions for the pilots, requiring them to turn around to read them. Additionally, emergency instruments were in multiple locations, which prevented pilots from getting all needed information with a single visual scan. The safety board recommended that the design and location of emergency gauges be reviewed (TSB, 2001).

In 1977, a system operator for the power control center in New York City accidently cut all power to the city when he turned a dial to the wrong setting. Lightning had taken down several electric lines, causing a power surge on the remaining lines. The operator was unaware of which lines were functional and not functional because his console did not adequately present information about overall system status. In a stressed state, the operator accidently turned a trip protection dial in the wrong direction—into the "Frequency Control" mode rather than the "Trip/Reclose" mode. Although these 2 settings were on the same dial, they had very different functions. Without trip protection, the remaining power lines over surged and Manhattan's entire power grid collapsed and could not be restored for 25 hours (Casey, 1993).

Several high-profile accidents have been attributed to inadequate displays and controls. The poor design of the controls in John Denver's Adrian Davis Long-EZ experimental aircraft contributed to the crash that took his life in 1997. The crash was determined to be due to the aircraft builder's decision to locate the unmarked fuel selector handle in a hard-to-access position; the fuel quantity sight gauges were also unmarked. Although partially due to the pilot's inexperience and lack of training, this accident may not have happened if the controls had not been so poorly designed (NTSB, 1998).

Bainbridge (2010) alludes to an issue with interface design from the video game domain. It is very common for Massively Multiplayer Online Role-Playing Games to include crafting skills. Players can work to create objects in the virtual world that can often be bought and sold in a virtual marketplace. However, the auction systems vary widely from game to game in their user interface, search tools, pricing and bidding policies, and other restrictions. This lack of consistency across games creates steep learning curves for new players and hinders players' ability to easily switch among games.

Perhaps the most famous event attributable to poor display and control design is the 1979 Three Mile Island nuclear power plant accident. Poor HCI nearly resulted in a nuclear disaster. Many of the controls and display system lights were poorly designed: the information necessary for operating the power plant was difficult to find, and the controls and lights conveyed either incorrect or confusing information to the user (Meshkati, 1991).

Robust, practical verification methods are required for displays and controls. Many standard measures of human performance exist, along with methods of data collection; however, some of these methods are not practical for use in an operational test environment. To address this gap, NASA has developed and researched verification requirements for legibility (Sándor & Holden, 2009), and usability error rate (Sándor et al., 2010b). Additional research is needed to develop and test integrated human performance metrics and verification requirements that include constructs such as usability and workload.

Exciting developments in virtual and AR and brain-computer interface (BCI) (Lalor et al., 2005; Lecuyer et al., 2008) present possibilities for future avenues of research regarding spaceflight. For instance, a quadriplegic individual was able to navigate a wheelchair along a virtual street using imagined movements of his feet (Leeb et al., 2007). Methods of interacting with AR (e.g. hand, pointer, and paddle) are also continually improving (e.g. Seo & Lee, 2013). As we begin to conceptualize LDEM habitats and mobile computing needs for these missions, we must be cognizant of new developments in human-computer interface technologies. Although some cutting edge innovations, such as BCI, may not be immediately applicable to habitat or mobile computing needs for LDEMs, it is possible that as they develop they can be adapted to fit the unique requirements imposed by spaceflight. Therefore, it is imperative that new innovations and research of human-computer interface technology be continuously monitored as we prepare for LDEMs.

Risk of Inadequate HSIA and Human, Automation, and Robotics Integration (HARI) Factors

Ground-based research that provides supporting evidence for the risk of inadequate HSIA and human, automation, and robot integration field is organized into 4 core contributing factors: (1) assignment of human and automation resources, (2) design for automation, (3) trust in automation and robotic agents, and (4) human-robotic coordination. This section also discusses methods for measuring human-system performance.

Human Automation Integration (HAI) Design

Assignment of Human and Automation Resources

The assignment of function or tasks to a specific agent (i.e., human, automation, or robot) can affect human performance. Different allocations lead to dramatically different performance levels. Inappropriate allocations lead to human operators reaching their performance and capability limits, increasing the risk for errors, injuries, and failed mission objectives.

The issue of assigning functions or tasks to humans, automation, or robotics has been around since the 1950s. Fitts (1951) proposed a set of human-computer role allocations based on the strengths and capabilities of each (summarized in Table 6). Fitts's list has been modified by many and used to develop other guidelines regarding task allocation and LOA. Because systems differ it is difficult to establish the tasks that should be automated and the degree of automation. de Winter and Dodou (2011) content that much in the human-automation world has changed since the inception of Fitts' list, and hence, it may be flawed, outdated, and incomplete, and that it disregards human-machine integration.

Table 6. Fitts's List (1951)

Humans are better at:	Computer are better at:
Perceiving patterns	Responding quickly to control tasks
Improvising and using flexible procedures	Repetitive and routine tasks
Recalling relevant facts at the appropriate time	Reasoning deductively
Reasoning inductively	Handling many complex tasks
Exercising judgment	Fast and accurate computations.

An alternative allocation list is the LOA, originally proposed by Sheridan and Verplank (1978). Parasuraman et al. (2000) expanded the list to include decision and action selection (Table 7, see also Sheridan & Parasuraman [2005]), and also proposed automation to be defined into 2 dimensions: *LOA* and *stages of automation*. Other frameworks to allocate human-automation functions are provided by Riley (1989) and Endsley and Kaber (1999).

Table 7. Levels of Automation (LOA), Parasuraman et al. (2000)

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

Scholtz et al., (2004) defines the following interactions based on human-robotic roles: supervisor interaction, operator interaction, mechanic interaction, peer interaction, and bystander interaction. Frameworks akin to LOA for human-robotic functional allocations have been proposed (Beer et al., 2014; Johnson et al., 2011). Beer et al. make a compelling argument for the adaptation of LOA to fit the robot autonomy. Fundamentally, the level of robot autonomy is based on the human and/or robot's role in sensing, planning, and acting (Table 8). The proposed framework may help designers to determine level of robot autonomy as a function of task, environment, and human-robotic interaction variables (e.g., SA, reliability, trust).

Table 8. Levels of Robot Autonomy, Beer et al, (2014)

Level of Robot Autonomy	Sense	Plan	Act
Manual	Human	Human	Human
Tele-operation	Human/Robot	Human	Human/Robot
Assisted Tele-operations	Human/Robot	Human	Human/Robot
Batch Processing	Human/Robot	Human	Robot
Decision Support	Human/Robot	Human/Robot	Robot
Shared Control with Human Initiative	Human/Robot	Human/Robot	Robot
Shared Control with Robot Initiative	Human/Robot	Human/Robot	Robot
Executive Control	Robot	Human/Robot	Robot
Supervisory Control	Human/Robot	Robot	Robot
Full Autonomy	Robot	Robot	Robot

In some ways, these frameworks for function allocation among human, automation, and robotic resources imply that a particular design operates at only one automation level. Opperman (1994) who describes adaptable or adaptive automation: flexible automation that can adjust and respond to user needs, environmental demands, and context. Under such circumstances, allocations between humans, and automation and robotic resources could change depending on the context of the work. Successful future space missions will require autonomous systems (e.g., unmanned vehicles, robots, rovers, habitats) that through human or system-induced modifications can adapt to unexpected events and endure extreme environmental conditions. De Visser et al. (2013) further discriminate between the adaptable and adaptive automation, describing the former as cases where change is induced by the operator and the latter where change is induced by the automated or robotic system. Parasuraman & Wilson, 2008 envision adaptive automation that adapts based on human state, which led to the emergence of neuroergonomics, where physiological measures are sensitive enough to support real-time, adaptive automation (Parasuraman & Wilson, 2008). For instance, neuroergonomics may lead to real-time assessment of operator workload and vigilance, which in turn could be used to determine when the operator is overworked or underworked and requires additional automation assistance. Neuroergonomics methods are common in research applications (Parasuraman, 2011) but not in real, safety-critical environments. Adaptive automation may be useful for LDEMs because of the many unknowns that exist. However, adoption of adaptive automation in human spaceflight operations is currently decades away.

LOA transitions also need to be considered for adaptive automation. Transitions can be sequential or discontinuous and go from fully manual to fully automatic or vice versa. Automation level transitions induce performance costs, including task inefficiencies, particularly during the engagement and disengagement processes. When an operator changes from one LOA to another, they must reorient to the current system status and operating level, which can decrease SA and precipitate out-of-the-loop performance costs. Di Nocera et al. (2005) maintain that the higher the LOA prior to LOA transition, the worse the "return-to-manual" performance. However, this may not be the case if the interface is designed to facilitate information sampling and thereby maintain operator SA (Di Nocera et al., 2005).

A meta-analysis of 18 published works on LOA transitions by Onnasch and colleagues (2014) indicates that automation necessitates a trade between LOA and human performance. Specifically, they analyzed the overall effect of the "degree of automation," which is the combination of *level* and *stages of automation* first proposed by Parasuraman et al. (2000). Although increasing the degree of automation helps in routine system performance and workload, it also negatively affects SA and operators' recovery from system failures (Onnasch et al., 2014) . The authors conclude that if return-to-manual performance issues are of serious concern, operators should be involved in the decision and action selection processes. Similarly, Hancock et al. (2011) advocates researching methods that would mitigate poor failure response when operating in higher LOA.

Much of the recent experimental research focuses on the LOA within a complex task and there is little evidence on the interaction of multiple systems with assorted LOA. For example, future astronauts may find themselves monitoring and commanding multiple, diverse automated systems that may be controlling life support, power, and communication systems, alongside robotic systems, such as rover scouts that explore new terrain and dexterous robotic arms. Calhoun & Draper (2015) examined the effect of different degrees of automation for a sequential primary task. The task involved managing multiple unmanned aerial vehicles, where the operator determined allocations (assignment of sensor tasks to vehicles) and routing (determining flight plans). Their findings suggest that if LOA is different across tasks, mode awareness issues may emerge, resulting in task performance variability. A NASA-funded research project, (Wickens et al., 2016; Christoffersen & Woods, 2002; Woods & Dekker, 2000) has examined attention allocations of the combination of 2 space-relevant tasks: telerobotics operations and a life support process control task. However, this work did not specifically examine the performance effects on assorted LOA among multiple systems.

Although LOA or robotic frameworks describe options for human, automation and robotic allocations, *determining* which of these allocations works best for a particular system is challenging. Often, human, automation, and robotic allocations are considered binary, one or the other, otherwise known as the substitution myth (Christoffersen & Woods, 2002; Woods & Dekker, 2000): the false assumption that if the functional allocation of a task switches from a human to an automated (or robotic) component, the resulting system remains unchanged.

However, different functional allocations between human, robots, and automation result in changes in operator work.

Limited guidance exists on how to apply allocations. Recently, the *Journal of Cognitive Engineering and Decision Making* devoted an entire issue to this very topic (Roth & Pritchett, 2017). As novel automation and autonomy emerge, research on human-automation interaction and the impact of allocations on human performance should not be limited to design and performance questions. Extensions or supplements to LOA may help "better anticipate and accommodate the variability of human response" (Roth & Pritchett, 2017, p. 4); an issue that is particularly relevant to human spaceflight which has a "small N" relative to other safety-critical domains (i.e., a finite number of astronauts).

Parasuraman et al. (2000) suggest guidelines rather than rigid rules are more practical to help determine tasks that should be automated and the LOA. They suggest first assessing the human performance consequences relative to the LOA to determine the range of appropriate LOA. The options can then be reviewed and narrowed down. Human performance consequences include factors such as safety, mental workload, SA, operator trust/distrust in automation, complacency, skill degradation, and mode awareness. Second, the reliability of automation and costs of decision or action should be evaluated. Methods to estimate reliability include fault and event-tree analysis and methods for software reliability analysis.

During spaceflight, task allocation for effective HARI is further challenged by how to appropriately determine the complete set of required tasks and functions because LDEMs never been accomplished before and the required technology may not yet exist. As a result, compiling a complete analysis of all tasks and functions is difficult. To guide the HSIA research plan, a basic list of the types of human, automation and robot tasks to be conducted in future LDEMs has been proposed (Marquez et al., 2017; Marquez et al., 2016); however, the resolution in describing these tasks remains high level (e.g., manage automated life support systems or drive surface rovers). Without adequately identifying the work functions (or tasks) needed, the partitioning of functions cannot be satisfactory. Thus, the effective allocation of function also depends on effective processes for needs analysis.

Typically, an analysis of all tasks and functions is called a needs analysis. This requires an understanding of mission operations, the particular functions that an automation or robotic system is supporting, and how these functions coordinate with related functions. Methods of need analysis include task analysis (Kirwan & Ainsworth, 1992), work domain analysis (Vicente, 1999), and contextual inquiry (Beyer & Holtzblatt, 1997; Woods et al., 2010). Furthermore, effective HARI design should to be integrated with process or workflow design (Beyer & Holtzblatt, 1997; Woods et al., 2010). Once an appropriate needs analysis has been conducted, this information must be linked to the design, development, and evaluation processes typically handled as part of software or systems engineering. The assurance that a system meets the identified needs is particularly important for safety-critical systems.

Although much research on function allocation addresses the principles for how functions should be allocated (Feigh & Pritchett, 2013; Hollnagel & Bye, 2000; Parasuraman et al., 2005), additional work is needed to provide the empirical base from which generalizations can be made. Shifts in recent thinking on generalized functional allocations is moving away from broadly defined, fixed "levels" of automation to emphasizing principles that are more fine-grained (Feigh & Pritchett, 2013; Hollnagel & Bye, 2000; Parasuraman et al., 2005):

- from component-wise allocation to integrated allocation
- from static allocation to dynamic, contextually specific allocation
- from function "ownership" by human or by machine to views where work functions can be shared or distributed across multiple human or engineered agents.

Function allocation methods, in turn, depend on methods for identifying and representing compatible work functions. Many methods that analyze existing work domains build detailed representations of task performance with a particular technology (and the function allocation provided with that technology), whereas approaches such as cognitive work analysis provide a more abstract representation in terms of system purposes. More methods are needed for efficiently representing work functions at the abstract or minimal level (Vicente, 1999; Pritchett et al., 2014a; Pritchett et al., 2014b). Abstract function specification will also be critical for novel applications where the work does not yet exist and no current work practice exists to study.

Pritchett and colleagues have proposed a function allocation method and corresponding evaluation metrics to assess various allocations of HAI (Ijtsma et al., 2017a; Ijtsma et al., 2017b; Ma et al., 2017), and they recommend a set of metrics that fall into 8 categories: workload, stability of the work environment, mismatches between responsibility and authority, incoherency in function allocations, interruptive automation, automation's boundary conditions, system cost and performance, and humans' ability to adapt to circumstances. A consistent set of metrics to evaluate function allocations and function allocation methods would enable NASA to develop human performance standards for integrating humans, automation, and robotics.

A variety of function allocation methods that are applicable throughout the lifecycle of system design are required, from preliminary concepts to defined technology requirements to real-time operations. Analysis of the work functions of the device and the human operator will be critical for effective work allocation and, hence, crew performance and mission success. It will be particularly challenging to develop a theoretical framework to guide how findings from one situation or task can be generalized to another. Furthermore, research is needed to understand how these frameworks must be adapted to include appropriate allocations in the context of robotics and real-time re-allocations. Integrated analysis is needed to determine how remote and on-Earth teams will collaborate with and through automated and robotic systems to achieve effective function allocation. Although methods for automation allocation exist, they do not translate to human-robotic allocations because the evaluation metrics for robots' capabilities are insufficiently described. Feigh and colleagues are extending their methods for automation function allocation to address the gaps with robotic allocations (Ijtsma et al., 2017a; Ijtsma et al., 2017b; Ma et al., 2017; Ma et al., 2018).

Design for Automation

As automation, robotics, and technology have become more sophisticated over the last few decades, so too has our knowledge of how design elements affect human performance. The research for this factor centers on attributes of automation, namely transparency of automation, and the many detrimental effects on human operators. Inappropriate LOA transparency leads to increased errors for human operators, injuries, and failed mission objectives.

Lack of automation transparency has been attributed to operators' inappropriate knowledge acquisition (Glover et al., 1997) and inability to maintain mode awareness (Sarter & Woods, 1994). Mode-related errors (i.e., misunderstanding the state of the system) are known contributors in aviation accidents and incidents. Two types of problems exist: difficulty in telling what mode a system is in and difficulty telling what a mode will do. For example, if pilot training teaches only a subset of possible modes, omitting unusual modes or those not used in airline operations policy, pilots can have latent problems in non-normal situations (Sarter & Woods, 1994). Problems can result because the system is in an unexpected or unrecognized mode or because the pilot could not predict behavior in an unfamiliar mode, or both (Sarter et al., 1997). In 1994, an A300 airplane crashed at Nagoya Airport. Prior to the crash, the mode had been inadvertently changed to "Go Around" rather than land. The pilot fought against the autopilot and was unable to gain control. Lack of awareness that the mode had been changed and inability to predict how the interaction between automation and human control would affect behavior in these conditions were both likely contributors (Sogame & Ladkin, 1996).

An Airbus flight test with a very skilled crew ended in a crash after a combination of conditions produced both an unexpected mode change and inability to compensate for the autopilot behavior in this mode (Aviation Week and Space Technology, 1995; also discussed by Sarter et al., 1997). This case involved an automation-initiated mode change due to the extreme conditions, and inadequate time to diagnose and recover from the control policy in this mode. Problems were compounded because automatic decluttering—designed to simplify the displays in an emergency—removed all flight mode enunciators from the display, leaving the pilot with no visible indicator of control mode. In another incident, unexpected effects of disengaging the autopilot led to pilot induced oscillations, resulting in passenger injury (Transportation Safety Board of Canada, 2008); lack of understanding the effects of mode change was a probable contributor. A similar incident has occurred with the Boeing 737 Max MCAS designs where the nose of the aircraft would pitch its nose down based on data from a single angle of attack sensor.

Sarter and Woods (1995) offer multiple solutions to alleviate the mode errors issue. These include reducing complexity of automated systems, finding new training methods and training in context, and developing new methods to facilitate mode awareness (e.g. displays that indicate current system state with justification and possible future states). Additionally, forcing functions could be introduced (to prevent a change until the current problem is corrected) or "management by consent" could be included in the system design (requiring operators to approve changes before they take effect). Nunes (2003) instead emphasizes development of accurate mental models as a method to support mode awareness and, hence, automation transparency.

Fundamentally, acquiring an accurate mental model of the automated system is critical for maintaining mode awareness. A mental model is the framework that manifests knowledge and understanding of a system's operating elements and processes to infer its future behavior. Mental models are formed from past experiences, expectations, and system feedback, and require time to develop. Nunes (2003) proposes the following strategy to train for accurate mental models based on the study of air traffic controllers using decision support tools: (1) Develop methods to determine what underlying mental model the operator has, and methods to assess the effects that automation has on human performance; (2) Develop tools that decrease cognitive workload while keeping information processing high, and train operators on how the automated system works and why it works the way it does; and (3) Educate users on the benefits and drawbacks of automation. Other researchers contend that decision support tools that decrease cognitive workload could jeopardize an operator's problem solving and learning ability because of the decreased cognitive effort taking place (Pea, 1993; Salomon, 1993; Wickens, 1992). This argument is partially supported by the Onnasch et al. (2014) who found that operator workload and manual skills are inversely correlated.

Mental models are essential for maintaining SA. SA is not only the perception of elements in current environment, but also the integration and comprehension of these elements, and the projection of future status based on comprehension (Endsley, 1995a). Dependency on automation can lead to decreased SA (e.g., Strauch, 1997). Endsley and Kiris (1995) concluded that higher LOA is associated with out-of-the-loop syndrome, the consequence of complacency and degradation in skill and SA resulting from prolonged supervisory control of automation. Kaber and Endsley (1997) showed that participants were better able to recover from automation failures when the LOA during the task involved human interaction (i.e., lower LOA). SA is essential for astronauts because they conduct tasks in an extreme, dangerous environment.

Studies of pilot-automation interaction within a glass cockpit aircraft (Sarter & Woods, 1995) assessed the relationship between mode awareness and automation complexity. Pilots were observed during their transition from a conventional to a glass cockpit. The pilot's lack of mode awareness and incomplete understanding of how the various modes interrelate was related to the complexity in the automation. Automation complexity, including simultaneous use of the system by multiple operators can increase the time to detect an anomaly and time to recover from errors. One pilot, for example, followed through correctly with a task because he was following standard procedure, not because he understood how the automation modes interacted. Another set of pilots had difficulty detecting an anomaly and/or deducing its consequences. Other pilots made changes to the automation but failed to activate them and did not understand why their changes did not take effect. Additionally, automation complexity can incite inadvertent mode changes, which may not bring about instant visible consequences (Sarter & Woods, 1995). To counteract this, indirect cues, e.g., prominent visible mode or goal indicators, could be provided by intelligent systems such as cockpit flight management systems (Feary et al., 1998). To achieve automation transparency, the operator must understand how the various modes interrelate, have a high SA of the current system state, past states, and possible future states, and be provided with appropriate feedback of system status and transitions.

Highly autonomous systems present a false dichotomy: the operator will either do too much or do nothing. Trades in design of automation can lead to appropriate human integration into these systems and adequate human performance. Furthermore, human spaceflight operations have taught us that astronauts must be kept “in the loop” for them to respond to anomalies and operate as backups. However, designing for transparency in complex, autonomous systems remains challenging. Johns & Kieras (1996) and Kieras (1997) describe 2 strategies used for handling complexity: distributing displays and controls across physical space, as in older “switch and meter” systems, and distributing displays and controls across time, as in newer software-based systems that use multiple modes. Information can be spread across time with different information displayed in the same space at different times, depending on context. In older electro-mechanical systems, displays and controls were fixed physical devices, and instrumentation was spread over the surfaces needed to contain it, leading to mode errors because information was often widely and awkwardly distributed. A critical factor in the Three Mile Island events was display layout: a distributed and awkward arrangement made the state of the system hard to grasp and failure diagnosis difficult (Leveson, 1995).

Transparency in automation and/or robotics interacts with functional task allocation. Human operators in charge of monitoring and potentially overriding an automated system may make or fail to make appropriate commands because they do not understand the system’s actual intention. Failures associated with misjudged intention are similar to simple mode errors because they occur when a mismatch exists between the operator’s estimate of the system’s intention (mode) and its actual intention (mode). Likewise, unexpected transitions from different LOA lead to operators being out-of-the-loop, which remains one of the biggest challenges in human-automation research (Endsley, 2017). Endsley and Kiris (1995) suggest the use of an intermediate LOA keeps the operator in-the-loop. Kessel and Wickens (1982) compared detection performance between operators controlling and monitoring system dynamics. Participants were placed in one of 3 groups: (1) transferred from manual operation to automatic, (2) transferred from automatic operation to manual, or (3) control group for automatic. Results from this study indicate that detection is slower when an operator is in an automatic mode than in a manual control mode. Additionally, a positive-transfer occurred from the manual mode to the automatic mode, implying that is it best to train operators in a manual LOA because they will be more perceptive of system changes once they transfer to a more automatic LOA. Having operators intermittently assume manual control may improve failure detection (Endsley & Kaber, 1999).

Transparency between automation or robotics and operators leads to mutual understanding of intent. Maintaining consistency between the operator and the system's intentions requires both active and reciprocal attempts to determine each other's current intent. This is particularly relevant for HRI. Very simple robots are capable of displaying spontaneous, recognizable human gestures (Cabibihan et al., 2012) and of appropriately interrupting ongoing inter-human communication (Saulnier et al, 2011) in ways that could facilitate communication between humans and robots. Conventional statistical or Bayesian methods have been used to infer users' movement intention during telerobotic operation from observations of their movements during test sessions (Dixon & Khosla, 2004; Ferguson et al., 2015). Operators can infer robotic system intent by observing indications as revealed by gaze direction (Wiese et al., 2013) or other social cues that can be directly provided by the robot's motion (Fiore et al., 2013).

During human-human team interactions, people seek predictive information and cues from each other. For instance, air traffic controllers have determined a number of visual cues predictive of pilot intent that they used to assist their planning of ground maneuvering at airports (Ellis & Liston, 2011). Similarly, this suggests HARI require predictive cues to infer intent. System intent may be displayed directly by explicit predictors that use visual projections to display future robot positions (Omidshafiei et al., 2015; Szafir et al., 2015). Feary et al. (1998) demonstrated the benefits of prediction aids in flight management systems.

Poor transparency concerning mode state and mode behaviors are most dangerous during unfamiliar conditions and operations in risky environments. Complex automation is powerful in part because it offers a choice of multiple control policies that can be selected based on circumstances. Nevertheless, managing different modes is key to effective integration. Understanding all the design trade-offs in managing complex control policies is essential to effective HARI. As such, NASA's standards and guidelines should be spaceflight-specific, extending known guidance such as Endsley (2017), to support human understanding of autonomous systems, minimize complexity of autonomous systems, and support SA.

Trust in Automation and Robotic Agents

People develop trust as they interact with automation and/or robotic systems. In turn, the amount of trust established has performance consequences. An additional complication arises as trust can change quickly and fluctuate over time. Achieving an appropriate level of operator trust in automation and/or robotics is imperative for safe and successful spaceflight. A recent internal trade analysis identified establishing appropriate trust in automatic and robotic tasks as an essential research area based on future LDEM needs and upcoming technological advancements in automation and robotics (Karasinski et al., 2019).

Since the seminal paper by Lee and Moray (1992), the human-automation interaction community has been studying trust in complex automated systems and the effect it has on human performance. The operator must be cautious not to over- or under-rely on automation or robotics. Over-reliance will diminish the operator's SA, placing all decision-making control in the hands of automation; under-reliance will inundate the operator's mental workload by relying solely on themselves. For example, operators did not detect the failure of a global positioning system component of the auto-navigation system on the Royal Majesty cruise ship. Their over-reliance on the automation resulted in the Royal Majesty running aground (Degani, 2004; Lee & See, 2004). Similarly, new research is focusing on understanding how trust in robotic agents affects human-system integration. Because robots are expected to transition from tools to teammates, the human-robot relationship will transition to a relationship that is similar to human-human teamwork where trust plays a critical role (Ososky et al., 2013a).

When trust is high, operators are likely to rely on automation, i.e., to act as indicated by alarms or hazard indicators, but may also be too "compliant" and fail to respond to unidentified hazards or hazards incorrectly identified by the system (Sheridan & Parasuraman, 2005). Over-reliance is also considered a symptom of automation bias (Cummings, 2004; Mosier et al., 1998; Skitka et al., 1999), i.e., when the operator tends to disregard or search for information contradictory to the automation, potentially resulting in errors of commission and omission (Mosier & Skitka, 1996; Skitka et al., 1999). Conversely, under-reliance emerges when the user does not trust the automation because it has failed too often, which could lead to automation not being employed (de Vries et al., 2003; Lee & Moray, 1994). Additionally, operator reliance on automation can be affected by workload (Kirlik, 1993; Wickens & Dixon, 2007), and by sudden change in automation performance (Lee & Seppelt, 2009).

Trust in automation and/or robotics is interwoven with functional task allocation. For example, issues may arise at the transition between manual control of a function to automated control because this change typically requires a reorganization of the operator's tasks (Sarter et al., 1997). Tasks that require automation, such as time-critical tasks, i.e., with little time for operator to response (Parasuraman et al., 2000), must be designed and built to be highly reliable to gain operators' trust.

According to Oleson et al. (2011), trust is affected by "trust antecedents," which can be classified as human-related, robot-related, or environment-related. Personality traits and prior experiences, which establish people's expectations, would fall under the human-related category. In the robot-related category, examples include consistency in performance and LOA. Shared SA and task complexity are traits that could be categorized as environment related. Oleson et al. (2011) acknowledge the importance of training to reduce biases and educate the operator on how the robot operates, including "capabilities…behaviors…and risks." Both trust antecedents and training, and any combination thereof, should be considered when designing a robot and during training.

Effective HARI design should be independent of reliability of either the automation or robotics because most spaceflight systems are one-of-a-kind and are likely to be modeled incompletely or imperfectly. This makes it particularly challenging to design effective integration between operators and systems. Consequently, it might be more useful to focus on understanding an operator's calibration with respect to the automation and/or robotic system in the context of trust. Calibration is the process by which the operator learns to modify their behavior based on the performance and reliability of the automated system (e.g. decision support tools, robotics) (McBridge & Morgan, 2010). For example, an operator's calibration may deal with how much time is spent on monitoring the automation, or how often the operator does what is recommended by the automation. The optimal calibration is not necessarily one in which the operator monitors or complies with the automation all the time—the operator may have other demands on their time and may justifiably believe automation is not always to be relied upon.

A complete and accurate mental model of the robot is important for appropriate calibration of trust. An inaccurate or incomplete mental model of the robot may lead to misuse and disuse (Hancock et al., 2011; Ososky et al., 2013b; Phillips et al., 2011). McGuirl and Sarter (2006) convey that miscalibration of trust results in a mismatch between perceived and actual system performance. Trust calibration is a multi-faceted process because appropriately calibrating trust in automation and decision support tools strongly affects mission outcomes. Methods to assist in calibrating trust include: (1) Fan et al.'s (1998) pilot decision-making study that revealed pilots initially rely on auxiliary information to validate an automated system's resolution, and with time, as system trust increases, the need for validation decreases; (2) Maes's (1994) interface agent model that illustrates how time affects trust, i.e., the longer the operator spends with an automated system, the more knowledge they gains about how the system works, thus, increasing their trust in the system; and (3) Cohen et al.'s (1998) model of trust in decision aids, which communicates that the operator must understand the automated system's strengths and weaknesses, and the context in which each prevail (McBridge & Morgan, 2010). A few instances where miscalibration can occur include situations when the automated systems' performance is unreliable, inconsistent, or not transparent, operator system knowledge is low, and operators have too much or too little confidence in their own abilities.

In a review of recent empirical research on factors that influence trust in automation, the 3 following interdependent layers of trust in automation emerged (Hoff & Bashir, 2015): (1) Dispositional trust is the most basic and stable trait and is influenced by culture, age, gender, and personality traits; (2) Situational trust depends on both the external environment (e.g., system complexity, potential risks and benefits of using automation, operator's workload) and internal human operator characteristics (e.g., self-confidence, mood, attentional capacity); and (3) Learned trust includes both initial learned trust (e.g., pre-existing knowledge) and dynamic learned trust. These observations suggested that an appropriate level of trust can be attained by providing operators with continuous feedback about the automation's reliability and situational factors affecting its performance.

Hancock et al. (2011) conducted a meta-analysis of factors affecting trust in human-robotic interaction, and categorized them into 3 areas: human-related, robot-related, and environmental. They found that robot characteristics, especially performance-based factors (e.g., reliability, predictability, transparency), played the most critical role in influencing trust, whereas environmental factors influenced trust only moderately, and little evidence existed of an effect of human characteristics on trust. Thus, the manipulation of the robot's performance and attributes may aid most in developing an appropriate calibration of trust. In a follow-on meta-analysis of a factors affecting trust in human-automation and HRI, Schaefer and colleagues did find human-related and automation-related factors influencing trust (Schaefer et al, 2016). Furthermore, they recommend further research to focus on the effects of human states, mode of communication, anthropomorphism, and agency transparency on trust development. Building on the effects of the robot's characteristics, a study conducted by DeSteno et al. (2012) showed that humans employ nonverbal cues to predict the trustworthiness of robots. Contrary to Hancock et al. (2011) and DeSteno et al. (2012), Desai et al. (2012) found that semantic association with risk and personal feelings about their own performance, as opposed to robot performance, were the strongest predictors of trust.

Systematic, objective measurement of trust in automation or robotic systems remains challenging. Currently, the most common method of assessing trust is through self-reported measures (Jian et al., 2000). Billings et al. (2012) contend that several questions must be addressed before quantifying trust. These questions include (1) What exactly is being measured? (2) How can trust be measured? and (3) When should trust be measured?

Although trust and transparency are integral in human-automation design and human performance, the outcome of such designs will ultimately be represented as a set of safety standards and guidelines. Crew safety is considered as part of verification and validation methods for HAI; human-robot proximity is a safety issue.

Human Robotic Integration Design

NASA currently has 3 robotic agents (Astrobee, Robonaut2 and SPHERES) on ISS that work in close proximity to the astronauts. Although they are technology demonstrations (i.e., they are not required to conduct operational, mission-critical tasks), their deployment raises important questions regarding how to make robots safe to work around humans. Rethink Robotics "Baxter^{TM}," a dexterous manufacturing robot, was built with a number of safety precautions such as torque sensing, collision detection, and power and speed restrictions (Anandan, 2013). This robot was designed for compliance, so no harm occurs to the human bystander operator if a collision happens. Current safety protocols are reactive, e.g., when a robot contacts a person it stops moving, a type of interaction that could have a negative impact on trust and mission performance. Further research is needed to understand the relationship between close proximity and human's perception of safety and trust. Prior work has shown that providing physical safety through collision avoidance is not sufficient to maintain human comfort (Meisner et al., 2008; Mumm & Mutlu, 2011; Takayama & Pantofaru, 2009).

Terrestrial research into robotic safety has gained traction as robots begin to appear in hospitals, nursing homes, and manufacturing facilities. The ISO is developing several safety requirements for collaborative robotics, e.g., ISO 13482 "Robots and robotic devices-Safety requirements for personal care robots." The American National Standards Institute has also recently approved the Robotic Industries Association R15.06-2012, "Industrial Robots and Robot Systems-Safety Requirements Standard." These standards are being developed to address safety concerns of humans working in close proximity to robots. The National Institute of Science and Technology (NIST) has a multi-year program in smart manufacturing, construction, and cyber-physical systems with the objective to "develop and deploy advances in measurement science to safely increase the versatility, autonomy, and rapid re-tasking of intelligent robots and automation technologies for smart manufacturing and cyber-physical systems applications" (NIST, 2014). Specifically, as it relates to human-robot collaboration, this effort is developing standards for safe operation of robots that are in close proximity to humans and determining metrics that measure how well a robot conforms to the standards.

Human-Robotic Coordination

Although similar issues exist for human-automation and human-robotic interaction, the latter has additional and unique challenges that affect human performance. Furthermore, in the context of human spaceflight, robotic systems have been necessity to meet mission objectives despite the increased complexity and risk associated with introducing these systems into this safety-critical domain. Well-established evidence shows that human-robotic coordination affects performance; however, with the advent of different, more capable robotic agents, HRI research continues to grow (Sheridan, 2016). The evidence summarized below is the most pertinent to human spaceflight with regards to the effects on human performance, safety, and mission success relative to human-robotic coordination.

Human teleoperation, supervisory control, and teaming have limited operational implementations in space, military, or industry settings. Currently, interaction with automation and robotics relevant to space exploration is achieved through teleoperations, i.e., one or more operators remotely control robot systems. During future space teleoperations, the operators and robots will be separated by large distances and a lag may negatively affect performance. At farther destinations, teleoperations (e.g., ground operators controlling Mars assets) will no longer be viable, forcing operations to switch to supervisory control. In addition, the LOA within the system will vary and the operator will at some point have to resort to supervisory control of the system, or robot(s). Supervisory control can cause degradation in SA and skill, system over-reliance or complacency, mode-related errors, and ultimately mission failure. Effective HARI must take these considerations into account when designing future HRI.

Many factors make the teleoperation of robots difficult for human operators. These can be categorized as procedural, geometric, and dynamic factors. *Procedural* factors include the semantics of the symbology associated with the robotic control, operational rules, and command sequences needed for specific tasks. Examples of *geometric* factors are the visibility of the worksite, the kinematics of the controlled elements, the reach envelopes and movement constraints of the controlled elements. Examples of the *dynamic* factors are the effective mass of the controlled element, its current velocity or other measures of its current physical state, and communication delays between the operator and robot.

The *geometric* and *dynamic* factors affecting telerobotics closely interact with each other in a manner that significantly affects the overall difficulty of teleoperation. The first example is the multiplicative interaction between difficulty and communication delay, i.e., system latency (Hoffmann, 1992). A second parallel example (Ellis & Adelstein, 2014) quantitatively demonstrates a nearly identical multiplicative interaction between rotational visual-manual misalignment and latency suggests that the cause of the similarity is associated with the operator and task rather than the specific cause of difficulty. The basis for the close quantitative similarity was argued (Ellis & Adelstein, 2014) to be the operator's default targeting process of successive reduction of the range error to the target in a classic intercept strategy called Pure Pursuit (Lin, 1991). The finding of a potentially generalizable and theoretically tractable multiplicative interaction between communication latency and control difficulty indicates a direction for developing principled guidelines for telerobotics subject to communication delays. Furthermore, the finding also suggests that human performance requirements for telerobotics can approach the quantitative rigor of those for manual control developed in the late 20th century (for review, see Jagacinski & Flach, 2003). The benefits of understanding the nonlinear effect of latency on telerobotic performance may be important because telerobotic dexterity may be affected by latency in an almost exponential way, and providing crew with low-latency telerobotic systems during LDMEs may be important (Geiger et al., 2010).

Li et al., (2014) evaluated the effect of increased automation specifically for the teleoperation of robotics arm. Using a test environment was arranged similarly to current space robotics workstations, they evaluated auto-guidance and auto-control of robotic arm. Consistent with previous research in aviation, increasing automation resulted in improved performance with lower workload, but poorer performance when encountering automation failures.

For efficient and effective remote navigation of a rover, a human operator needs to be aware of the robot's environment (Nielsen et al., 2007). During teleoperation, operators receive information about the environment only through a robot's front-mounted camera. Receiving limited visual cues from a single camera view causes a so-called "keyhole effect" (Voshell et al., 2005), which reduces the operator's SA and may result in navigation issues such as more collisions, not noticing critical aspects of the environment, or reduced speed (Casper & Murphy, 2003; Tittle et al., 2002). The configuration of cameras on or near a robot can have significant effect on the number of operator input errors and mental demands during teleoperation.

Research on camera configurations for space robot arms have consistently demonstrated that the less mental or perceptual compensation the human performs to associate input with output, the better task performance will be (Bray et al., 2003; DeJong et al., 2004; Menchaca-Brandan et al., 2007; Smith & Stuart, 1989). To compensate for visual ambiguities in the video feed from a robot, portions of the robot's chassis can be included within the camera's field of view to provide operators information about the size of the robot relative to objects within its environment. This may lead to better obstacle avoidance and enhanced ability to navigate at faster speeds. Additionally, operators may be able to infer the attitude of the robot (i.e., pitch, yaw, and roll) and thereby preventing commands that would cause the robot to roll or flip over (Wang et al., 2004).

Multiple camera views appear to help operators maintain awareness of a robot's surroundings (Scholtz et al., 2004). Visual cues from multiple cameras (along with the inclusion of the robot chassis, as noted above) further compensate for the keyhole effect and ambiguities operators experience when they teleoperate a robot (Wang et al., 2004). The frame of reference of cameras also influences operator performance. In the exocentric (i.e., third-person perspective) frame of reference, the human operator views the robot's environment from a fixed location external to the robot's workspace. In the Multi-Mission Space Exploration Vehicle, an exocentric view may be achieved by placing cameras outside of the vehicle on a lunar or planetary surface, although doing so may not always be feasible. In the egocentric (i.e., first-person perspective) frame of reference, the operator views the robot's environment from the perspective of the robot.

For a teleoperated driving task, the operator is often provided with an egocentric (i.e., first-person) view from a camera placed on the front of the robot (Keyes et al., 2010). Although an egocentric frame of reference may limit the mental rotations and translations the operator needs to perform to move the robot (e.g., rotating a joystick makes the robot rotate in the same direction), operators given this frame of reference may lose spatial awareness of the environment around the robot (Chadwick et al., 2006). For example, operators may have inaccurate perceptions about the size of objects and the distance from the robot to those objects (Casper & Murphy, 2003). The addition of an exocentric (i.e., third-person) field of view from a camera placed in the robot's environment may provide operators with the additional information needed to gain SA of the robot.

Displays can be modified in various ways to improve human-robot interaction. Improvement in perception allows for better understanding of the environment, improved function (Chen & Joyner, 2009; Darken & Cervik, 1999), and decreased operator workload (Park & Woldstad, 2000; Yeh & Wickens, 2001). Optimal views are likely to be determined for the specific task (Smyth et al., 2001). Additionally, research suggests that a third-person point of view during HRI (which is referred to gravity-based orientation) improves performance (Thomas & Wickens, 2000), and that auditory and tactile feedback can improve operator performance when task complexity increases. Prewett et al. (2010) propose various ways of improving operator performance, while decreasing operator workload. These include (1) increasing the camera frame rate to an optimal human information processing level, (2) minimizing system delays, and (3) providing naturalistic displays. Pazuchanics (2006) suggests incorporating contextual information into interface features can decrease operator workload.

Currie and Rochlis (2004) conducted a study using the SPDM trainer to assess the feasibility of ground control operations under time delay, a condition that will become of primary concern for robotic operation during LDEMs. Astronaut test subjects conducted an Orbital Replacement Unit Removal and Replacement activity with simulated 6 and 8 second telemetry and video time delays (loss of signal, LOS, conditions were not modeled in this experiment). Crew found that they could adapt to the delays and all subjects were able to successfully complete the activity. Interestingly, more than 50% of the ground control operational time was spent not on motion command inputs via the hand controllers, but on manipulating displays, cameras, and controls to gain and maintain SA, which implies that performance increases can be gained from more effective interface designs (Currie & Rochlis, 2004). Other NASA-funded studies have examined the effect of low-latency on robotic arm manipulation, but the human performance results have yet to be published.

Human interface enhancements such as AR have also been investigated for SPDM (Maida et al., 2007). Although a HITL command mode is available for SPDM, it is operated almost exclusively using automated sequencing from the ground, where camera views and SA information is minimal. Results indicate that overlays improve performance in maneuvering an Orbital Replacement Unit and in prepositioning accuracy in preparation for inserting it into a receptacle.

The design of controllers is another aspect of human interfaces that should be considered as robotic agents become more dexterous. Currently, the state-of-the-art for operationally proven space robotic dexterity is SPDM. Increased dexterity in robotic agents will be necessary for a variety of future NASA missions. In 2005, the NRC concluded that a robotic service mission for Hubble Telescope was too risky (NRC, 2005). It cited "current low level of technology maturity" among its reasons, which included the yet to be deployed SPDM and dexterous manipulators robotically operated from Earth. Other advances in dexterous robotics includes work at NASA Goddard Spaceflight Center (2015) on a series of robotic refueling demonstrations aimed at advancing satellite maintenance. The European Space Agency has emphasized human interface design in developing and evaluating telerobotic interfaces and haptic controllers (e.g., Aiple & Schiele, 2013; Heikkilä et al., 2012).

Future HSIA-Robotic designs must include dynamically responsive dexterous robotic systems (comparable to the Da Vinci telerobot (Hubens et al., 2003)) to complete tasks ranging from science exploration to repetitive maintenance. Recently during a Deep Space Gateway Science Workshop, scientists envisioned an exploration campaign that significantly leveraged humans, robots, and human-robot teams (Da-Poian & Koryanov, 2018; Head et al., 2018). Such high dexterity will be novel in NASA's spaceflight operations, and human performance standards must be adapted from the research already being conducted in the telerobotics medical field. For instance, Anvari and colleagues (2005) suggest that latencies beyond 500 ms will result in deteriorated performance when using a telerobot to conduct a surgical task.

Although there are many aspects of human-to-human interaction, a desired characteristic of human-robotic interaction is the human's ability to apply alternative viewpoints and reason from the alternative viewpoints. Trafton et al. (2005) assessed how perspective-taking is already used or could be applied to spaceflight EVA by observing and analyzing EVA training activities at the NASA JSC NBL, and they determined that spatial perspective-taking was occurring during these activities. For example, challenging spatial perspective-taking includes interpreting of the word "down" relative to a robotic end-effector with 6-DOF and in microgravity. The 3 guidelines for perspective-taking include: (1) ensuring that all aspects of robotic representation and "cognitive" functioning (such as reasoning and perception) are as human-like as possible, (2) building cognitive systems for the interactions based on integrated cognitive architectures, and (3) applying heuristics and principles for collaborative activities that align with human expectations. Overall, similar to human-to-human perspective taking, a robot should be able to assume and adapt to multiple perspectives while performing tasks to allow for optimized human-robotic interaction. Trafton et al. discussed difficulties associated with studying collaboration between humans and robots, and they attempted to answer questions such as when to collaborate, when to ask for help, and how to respond to assistance.

A new area of human-robotic interaction is the supervisory control of many robots by one operator. Swarms of robots imply a large group (or groups) of simple robots that perform a collective task beyond the capability of a single robot. Although not yet integrated into future human spaceflight mission designs, these robotics (e.g., Swarm Orbital Dynamics Advisor or HelioSwarm [Plice et al., 2019]) are being envisioned to support human planetary exploration. Kolling et al. (2016) present a survey of human-robotic interaction challenges with robot swarms, which includes human-swarm communication, state estimation and visualizations, and human command and control of robots.

Research on human-robotic coordination has centered mostly on commanding and controlling robotic arms or rovers, be it the effect of HRI design, mission constraints (e.g., inclusion of latency), or human's cognitive capabilities (e.g., spatial ability). Future missions will include with human-robot teams (Fong & Nourbakhsh, 2005). that will operate jointly, share goals and intent, and deal with issues regarding collaboration, cooperation, and communication. As such, there is a new emphasis on HRI focused on team performance (Ma et al., 2018).

To build a successful human team, team members must share a common goal, share mental models, suppress individual needs for group needs, view affinity as positive, understand and achieve their role within the team, and have trust for each other (Groom & Nass, 2007). Robots, however, do not have mental models, individual values and beliefs to guide them, or even self-motivation, and, therefore, lack the essential qualities of a successful teammate. It is difficult to establish fixed criteria for operating teams of humans and robots because each human will bring their own prior experiences and knowledge, and the allocation of tasks between humans and robots will differ from mission to mission. To mitigate these issues, Groom and Nass (2007) developed fundamental questions that need to be addressed to determine operating policies for human-robot teams:

- What are the restrictions of humanness that limit performance?
- Which inabilities of humans can be successfully implemented in robots?
- What organizational structure best optimizes both human and robot capabilities?
- If a particular organizational structure is ideal for humans, do robots have the potential to fulfill the social duties of the role, and is developing those abilities valuable?

One essential component to successful human-robot teaming is communication. Traditionally, command and control of robots was envisioned as human-robot communication. In a team context, however, communication is about cooperation and collaboration. HRI research is currently focused on understanding human intent, which includes subtopics such as robotic interpretation of human intent, understanding intent unobtrusively, and improving human-robot communication. Computers or robots can interpret human intent in a variety of ways, including speech, gestures, and other forms of nonverbal communication. These techniques are at varied levels of development, from basic proof of concept to use in operations. Each technique can be broken down into several levels—gestures, for example, have 4 levels: sensor technologies, identification, tracking, and classification (Liu & Wang, 2018). Research is also focusing on interpreting behavioral and human monitoring data and integrating the data into human behavioral models to predict intent. For instance, eye movements are being studied to predict human intent (Ruhland et al., 2015; Singh et al., 2018). A recent internal trade analysis identified methods for understanding human intent for HARI tasks as an essential research area based on future LDEM needs and upcoming technological advancements in automation and robotics (Karasinski et al., 2019).

In human-human teaming, shared mental models (SMM) provide a long-lasting, common understanding of task knowledge, processes, and team dynamics. Shared SA represents the degree to which the team shares an understanding of the current operational picture and is influenced by the SMM. These lead to effective teams being able to understand and interpret a situation in the same way, know what the other team members already know, know what they need to know, and know what they are doing (Cooke et al., 2000; Nikolaidis & Shah, 2013). Promoting robots from useful tools to synergistic human-robot team members will require applying SMM and shared SA to mixed human-robot teams, particularly when the team must adapt to unexpected situations. Although numerous studies have modeled the performance-linked characteristics of SMMs and shared SA in human-human team coordination (for review, see Mohammed et al., 2010; Saner et al., 2009), efforts to applying these models to a human-robot interaction framework ae limited.

SMM frameworks for human-robot teams have been proposed, however, these frameworks have been difficult to apply into cognitive robot architectures (Adams et al., 2014). SMMs depend on successful, bi-directional communication between humans and robots, yet research indicates that this communication is affected the way humans perceive robots. Different hardware or physical components of a robot elicit changes in SMM development (Hwang et al., 2005; Kiesler & Goetz, 2002). People are more willing to interact with robots that have human-like features, and subsequently, form a richer mental model (Kiesler & Goetz, 2002). Ososky et al. (2013b) report that people fell they have more knowledge of a robot with anthropomorphic traits than those without, suggesting that humans tend to interpret a robot's knowledge and ability just as they would another human (Lee et al., 2005).

A few studies have started to investigate the effect of human performance in the human-robot team. Recently, Gervits et al. (2018) studied SMM for human-robot teaming in the context of spaceflight by focusing on the free-flyer robots (like Astrobee). Preliminary results indicate that teams with SSM could complete tasks more efficiently, however, there was no reduction in workload or SA. Other investigators have studied mixed-initiative operations: an interaction strategy where the best-suited member of the team performs (Jiang, 2019; Jiang & Odom, 2018). For example, Losey and colleagues recently investigated intent detection and arbitration of robot control between human and automated systems (Losey et al., 2018). The emerging human factors issues that impact human performance are summarized by Chen & Barnes (2014) for human-robot teaming:

- flexible human-agent interaction
- maintaining operator's ultimate decision authority
- supporting operator's multitasking performance
- enabling automation transparency
- visualization and training techniques enhancing human-agent collaboration
- incorporating human individual differences in design process.

Systematic assessment of human-robotic interaction and the methods and metrics to do so are still emerging (Ezer et al., 2013). Additional factors to consider (Kadous et al., 2006; Scholtz et al., 2004) include awareness, efficiency, familiarity, and responsiveness of HRI. Awareness involves the proper presentation of information to ensure operators have a complete mental model regarding the robot's internal and external state, which avoids overloading the operator with too much information. Efficiency involves requiring as little operator hand and eye movement as possible and ensuring focus of attention. Familiarity focuses on including intuitive information and concepts that the operator is familiar with and avoiding the presentation of unfamiliar information. Finally, responsiveness guarantees an operator is receiving continuous feedback regarding operations.

Ma et al., 2018 contend that team-centered metrics beyond system, operator, and robot performance help quantify collaboration, cooperation, and coordination. Olsen and Goodrich (2003) contend that the goal of human-robot interactions is to minimize the time an operator spends manipulating a robot (robot attention demand or RAD), allowing them more time to focus on other tasks. RAD is derived from the following equation: RAD = IE / (IE + NT), where IE (interaction effort) is the amount of time required to interact with a robot and NT (neglect tolerance) is the robot's autonomy. The difficulty lies in determining IE because interaction may be cognitive, which is not as readily apparent as physical manipulation (Olsen & Goodrich, 2003). Other variables that must be considered when determining operating policies for teams of humans and robots include the human's trust in the robot, task context and complexity, the number of operators and robots, interface design, and goals.

Keyes et al. (2010) suggest that overall awareness is key to successfully completing robotics tasks. Robotic operations often occur outside of the human operator's direct line of sight, requiring the human to have acute knowledge of all aspects of the working conditions and environment that the robot is operating and the state of the robot (Keyes et al., 2010). Five levels of human-robot awareness include human-robot awareness, human-human awareness, robot-human awareness, robot-robot awareness, and the human's overall awareness of the mission (Drury et al., 2007). Kadous et al. (2006) contend operators must relate and/or sense the robot's operations to achieve satisfactory SA through human-robotic interaction.

User experience may be critical in the effectiveness of human-robot collaboration. Preliminary investigations of various NASA robotic systems (SSRMS/SPDM, R2, Curiosity, and SPHERES) indicate that pragmatic qualities (task related goals such as how quickly is the task performed) are more important than hedonic qualities (non-task related goals such as enjoyment performing the task). Further research may provide evidence for the relationship between user experience dimensions, operational context, and operator roles.

Ferketic et al. (2006) highlight the increasing need to establish standards for defining effective human-robotic interaction and designing interfaces. Although human engineering and HCD standards can apply, specific guidelines are needed due to the unique demands of LDEMs and the nature of the interactions and objectives related to human-robotic operations. Developing HRI standards can be extremely complex because most robots are developed with customized interfaces and methods of interaction, and the level of coordination and control required is often highly task dependent. Common metrics and measures must be developed to standardize different types of operations and interactions, including fundamental commands, operations, and interfaces that lead to expected and similar responses from different robots. This will subsequently lead to increased consistency in robotic actions, operators' familiarization and comfort with control and operation of the robots, and decreased risk of errors and mission objective failures. Ferketic et al. (2006) proposes the following guidelines that will allow for the evolution of human-robotic interaction, increased safety, and successful, collaborative task completion during spaceflight:

- define the capabilities and limitations of humans and robots
- develop user interfaces that are suited to the task at hand

- address any related challenges to efficiency and operations to allow for human-robotic collaboration
- establish superior prototyping and evaluation methods.

Human-system performance must be appropriately quantified to recommend HSIA and human-automation and robotic guidelines and standards for spaceflight operations. However, developing *operationally relevant* guidelines and standards will be challenging. Although significant ground-based evidence exists suggesting the appropriate metrics, limited spaceflight data is available to determine operational limits. For instance, it may only take a couple of seconds to detect automation failure in a lab, yet in an operational setting, detecting a non-critical automation failure in the order of minutes might be acceptable. The challenge becomes generalizing research metrics and limits in an operational setting.

Some data from spaceflight operations are retained and stored for posterity (for example, NASA records all voice communications between ISS and ground controllers), however, no *systematic* collection of data on task performance has been conducted during spaceflight. The Human Performance Database Project is focused on identifying how data on operational task performance could be retained to inform the HSIA risk. The Human Performance Database Project used spaceflight robotics operations as a case study, and identified subjective data on training measures and relevant ISS telemetry that could be captured real-time as part of robotics task performance. Key to their case study is the need to also capture operational "context", e.g., dates of tasks, call downs back to ground, and camera selection. Future efforts are underway to improve systematic operational data collection.

Many human performance behavioral and physiological measures and metrics have been used in ground-based research to describe human attributes, capabilities, and performance quantitatively and qualitatively. *Measures* are amounts of a phenomenon that can be directly quantified in terms of absolute or relative physical units (e.g., relating to fundamental units of mass, length, time, etc.) collected by specialized instruments, by directly counting the number occurrences, or by rating scales or preference rankings elicited from human observers. Phenomena or items that can be measured encompass not only the output of engineering systems, processes, and components, including their human operators, but also their inner workings (i.e., parameter settings and input-output relationships). *Metrics* are values that are not necessarily measurable directly, but are derived from one or more measures to indicate the degree to which a phenomenon exhibits a given attribute or construct, and metrics afford a means for comparison with benchmarks (Blanton & Jaccard, 2006). Thus, measures and metrics serve to describe (i.e., model) and control the operation of systems. Measures and metrics, therefore, also enable the effectiveness and efficiency (i.e., performance) of systems and their operators to be assessed for purposes of basic and applied research, design iteration, requirement verification, product acceptance, and maintaining and improving productivity.

Behavioral quantities that describe human actions and those of robot hardware and software systems can be measured objectively. Typically, these measurements result from the states of hardware sensors, clocks, and actuators, and from software state information that monitor and drive the physical actions of the robot. For example, simple events such as the success or failure of target acquisitions by the human robot system over the course of an activity can be counted. Likewise, the robot's sensors can measure the distance of the robot end-effector from a target. Accumulated over the course of a number of trials or a given time interval, these simple raw counts and displacement measurements can be converted to error rates and average displacement errors, serving as metrics that may help quantify the effectiveness or efficiency of robotic task. Moreover, measurements of human inputs (e.g., manual displacements and forces and verbal communication) and multi-dimensional robot kinematic and dynamic outputs can be combined in more elaborate ways to yield more nuanced metrics for a human-robot system.

Well-known methods that elicit subjective scale ratings from operators to assess handling qualities (e.g., Cooper & Harper, 1969), workload (e.g., NASA-TLX [Hart & Staveland, 1988]; Bedford [Roscoe & Ellis, 1990]), and SA (e.g., SAGAT [Endsley, 1995b]) can be used to assess human performance with complex engineering systems. Elicited rating, however, must interrupt operators at discrete test break points or at the end of a task. Alternatively, the human operator's spare capacity, an indicator of workload, can be estimated through objective performance measures collected from concurrently executed secondary tasks without interrupting the main task. Physiological measurements such as heart and respiration rate or galvanic skin response, which can correlate with the psychological state (e.g., stress or fatigue), may also help elucidate performance factors such as workload or handling qualities. Such physiological monitoring, when sufficiently nonintrusive and once validated, could be suitable for HSIA-Robotic evaluation. Additionally, NASA-funded research is evaluating real-time measures of task performance without interrupting operators (Duda et al., 2015); specifically, validating HSIA-Robotic metrics that include flight performance, SA, and workload for simulated tasks ranging docking operations to EVA jet pack self-rescue.

NASA-funded research on human-automation interaction is developing a framework to combine a variety of measures and metrics of human-system performance, safety, and efficiency (Oglesby et al., 2014; Stowers et al., 2017). Performance determines whether task goals are achieved with the given system, whereas efficiency determines whether that performance was accomplished with minimal expenditure of time and/or effort. Safety assesses injuries or harmful events that arise from interacting with the complex, automated systems.

The type of robot (e.g., manipulator, rover, or free-flyer) dictates the specific types of measurements and metrics used to assess human-robotic interaction. Many metrics have been described in the literature and still more can be devised. For example, De Barros and Lindeman (2008) describe 48 HRI metrics of high-level tasks (e.g., manipulation, social interaction, navigation, perception, and resource management) that could be performed by mobile robots (Steinfeld et al., 2006). Although some of these metrics have not yet been fully defined or implemented, many are based on objective measurement and others are potentially subjective in nature. Similarly, Forman (2011) lists 33 objective metrics of manipulator-arms and 3 objective metrics of cameras that have already been used in the Massachusetts Institute of Technology Man Vehicle Laboratory simulator for telerobotics research, and the author notes the lab's capability to investigate many more.

Other HRI metrics that have been proposed range from simple measures of inherent system communication latency (time delay) to more advanced constructs such as neglect, i.e., how long an operator can attend to other matters while a robot runs autonomously (Crandall & Goodrich, 2001), and fan-out, i.e., how many robotic agents can one operator simultaneously control, which is a function of neglect (Olsen Jr. & Wood, 2004). Ultimately, these metrics are intimately related to the autonomy of robotic systems (Crandall & Goodrich, 2001; De Barros & Lindeman, 2008; Ong et al., 2008; Yanco & Drury, 2004) and affect system usability (Weiss et al., 2009).

Measurable, valid, and reliable metrics are critical to support HSIA-Robotic design, and to objectively evaluate and contrast the relative merits of the design (Blanton & Jaccard, 2006; Carmines & Zeller, 1979; Weiss et al., 2009). Methods and metrics must be validated to demonstrate that they are sufficiently robust and useful (i.e., stable, sensitive and reproducible) to support development of knowledge bases, theoretical frameworks, and the consequent models necessary for the rational design and evaluation (e.g., requirement verification) of a broad range of human-systems (Dautenhahn, 2007). Although researchers and system developers have adopted already existing metrics and methods from other domains such as HCI, human factors, and psychology, these too must be validated before they can effectively serve in an HSIA-Robotic context (Fong et al., 2005).

Risk of Inadequate HSIA and Training Factor

The ground-based evidence presented here includes controlled studies (Evidence Category I or II), case studies (Evidence Category III), and expert opinion (Evidence Category IV).

Foundational Principles of Training

Ground-Based Evidence: Retention

Sanli and Carnahan (2018) reviewed literature on skill retention and multi-day training in various fields (medical, military, marine and offshore safety), and found that skills and knowledge of Advanced Life Support and CPR appeared to decline between 6 months to one year after training, although information was limited during the first 6 months after training so skills may decline more rapidly. The authors report that skills appeared to decline faster than knowledge, as confirmed in a different population (neonatal resuscitation by medical staff).

Sanli and Carnahan (2018) reported the decay rate for soldiers' basic training skills was associated with the number of steps in a given skill, the presence or absence of subtasks within a given skill, and the serial order (functional task difficulty) in which the tasks were trained. The more steps (nominal task difficulty) that were involved in the task, the more difficult the task. Functional task difficulty can be reduced by enabling learners to review notes, or use a technical manual for reference, as described in another study by Sanli et al. (2018) who reported that solders retained complex skills at a high level for 4 months. Tasks that required a shorter time to complete (less steps), had less interference (similar or same tasks performed in a similar manner), and were practiced more often were better retained than other tasks. Nearly all skill components decayed across retention intervals, but practical skills were found to decay more quickly than other types of knowledge. Although both long- and short-term accumulation of practice influenced retention performance, Sanli and Carnahan (2018) contend that some tasks, especially complex tasks that involve movement skills, can be retained for at least 6 months. Sanli and Carnahan (2018) describe nominal task difficulty as inherent task difficulty, regardless of who is completing the task or under what conditions it is being performed; whereas, functional task difficulty reflects the experience of the learner and the conditions in which a task is performed. The authors contend that retention performance is affected by functional task difficulty; for example, what order a task is performed and whether a manual can be referenced for guidance during a task affected retention.

Siu et al. (2016) investigated an adaptive VR surgical training system (i.e., peg transfer task and needle passing task for VR laparoscopic surgery) for U.S. Army medical physicians, to track learning, retention, deficiencies, and generate an optimal adaptive training schedule. The researchers assessed decay of surgical skills and they developed training methods to prevent decay, minimize training time, reduce mistakes, and maximize efficacy. Performance of novices and medical trainee on these tasks (peg transfer and needle passing) were assessed at 3 sessions: at baseline, 1 week after baseline, and 1 month after baseline. An adaptive control of thought/rational cognitive architecture was used to model the learner's learning and forgetting, and 3 different stages of learning and forgetting were proposed (Figure 36).

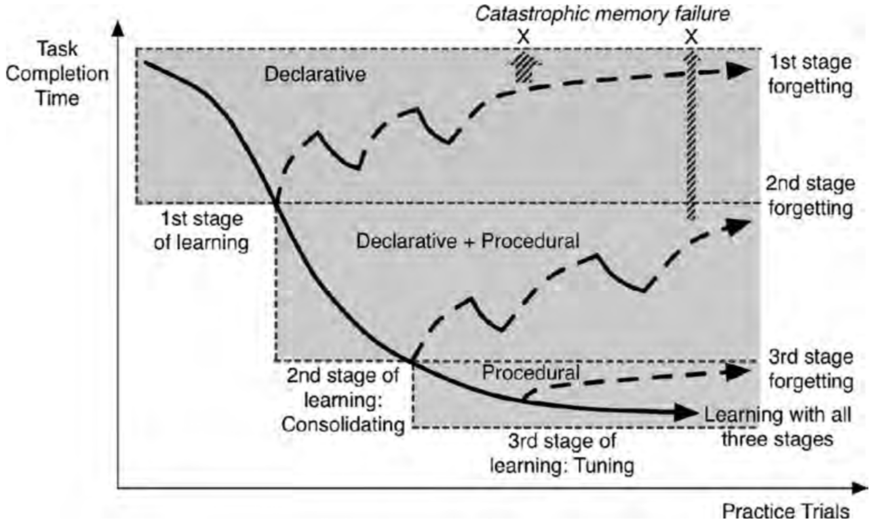


Figure 36. A Theory of the Stages of Learning and Forgetting. The solid line shows the learning curve during each stage, the dashed line shows the forgetting curve (Siu et al., 2016)

Completion time increased during skill acquisition (during the first session), then slowed across sessions during periods of disuse. The skill decayed less if practiced more. The patterns of performance were better predicted by the Power Laws of Learning and Forgetting, within the adaptive control of thought/rational model, than a logarithmic regression model.

These and other studies provide clear evidence that training in medical skills, military skills (Sanli & Carnahan, 2018) and surgical skills (Siu et al., 2016) decay during intervals much shorter (~ months) than a planned mission to Mars (~ years). Research is needed to quantify the skill decay of critical spaceflight mission tasks in the same manner as studies conducted on terrestrial based tasks. The retention intervals determined from this research can then be used to provide a program of training that supports ongoing skill maintenance.

Ground-Based Evidence: Transfer

Learning is highly specific to the conditions under which it occurred, especially when the learning involves procedural, as opposed to declarative, information. To account for this specificity, Healy and Bourne (1995; see also Healy, 2007; Healy et al., 2005) proposed a procedural reinstatement principle, i.e., training on one skill does not transfer to another related skill unless the procedures required by the 2 skills overlap. This principle is related to other principles and theories in the literature, including Thorndike's (1906) theory of identical elements (see also Rickard & Bourne, 1995, 1996; Singley & Anderson, 1989); Tulving and Thomson's (1973) encoding specificity principle; Morris et al. (1977) transfer-appropriate processing principle (see also McDaniel et al, 1978; Roediger et al., 1989); Proteau et al.'s

(1992) specificity of practice theory involving sensorimotor representations for motor learning; and Kolers and Roediger's (1984) theory involving procedures of the mind. These empirically driven principles and theories drive current research to develop and design spaceflight training programs (Barshi, 2015).

Training aimed at long-term retention often results in high specificity, and thus low generalizability (e.g., Vogel & Thompson, 1995). Specificity might be desired when all tasks are well known and understood in advance; however, that cannot be the case for future LDEMs. Because LDEM crews will face situations and tasks are currently unforeseeable, future training should focus on generalizability of skills, rather than specificity. Current research has made important, albeit small, steps towards understanding generalizability, but shifting that understanding to practicable training principles has proved challenging (e.g., Healy & Bourne, 2012). Billman and Catrambone (2019) investigated an instructional principle of "providing and integrating experience with device models and with operational procedures" but did not find the strength of generalizability that they had predicted; thus, a clear method for effectively training generalizable skills is still far off.

Ground-Based Evidence: Simulation Fidelity

Marcano et al. (2019) reviewed simulator training practices in different industries and determined that some individual simulator training can supplement traditional training practices, that proper strategies should be the basis of individual simulator training, that real-time feedback and assessment is important, and that individual simulator training should include an embedded intelligent training system. This review of training practices can be very informative in the design of spaceflight training; however, designing training simulators for critical mission tasks may require more detailed insight.

Many aviation studies investigating the effects of variations in motion fidelity on pilot behavior and performance (Schroeder & Grant, 2010; Zaal et al., 2015) have shown that motion cues strongly affect pilot skill-based behavior and performance of some tasks. Motion cues are more important in instances that require human-led equalization (i.e. a response to rate information); for example, when controlling unstable aircraft dynamics. Furthermore, motion cues are the greatest in motion error-reduction tasks, where the main visual information is the error between the desired and actual state of the vehicle. When the task is more of a disturbance-rejection task, motion helps to increase performance, whereas during a target-following task, motion helps to increase stability margins. Motion can also improve performance in conditions with higher workload, such as when multiple axes (e.g. pitch and roll) are controlled at the same time.

One of the main problems with investigating the effects of simulation fidelity on human performance and training is the fact that no standardized metrics for fidelity exist. Rather than comparing the response of the simulated vehicle dynamics and systems, including the stimuli generated by the simulator cueing hardware and software, to responses of the actual vehicle, human-centered fidelity metrics have been proposed. These metrics consider human perceptual limitations; i.e., that reduced fidelity in certain areas might not be perceivable. The most basic human-centered fidelity metric is simply asking a human operator to rate the fidelity on a scale. However, human responses are highly variable and depend on many factors such as fatigue and boredom. Other human-centered fidelity metrics have focused on modeling human perception processes and human behavior in system-theoretic models (Mulder et al., 2013).

Research studies show both the importance of providing simulator training to trainees and the complexity of developing effective simulators, whereas accident investigations show the tragic consequences of not providing effective simulator training. In a Congressional Report investigating the 2 accidents on Boeing's 737 Max aircraft that resulted in the deaths of 346 people, the committee stated that while the captain and first-officer on the Ethiopian Airlines flight were seasoned pilots, "the 29-year-old captain of Ethiopian Airlines flight 302 had reportedly not received training on the airline's 737 MAX simulator—although Ethiopian Airlines was one of the first airlines worldwide to purchase a 737 MAX specific simulator." (Defazio & Larsen, 2020, p.10). A significant finding of the report is that, "In January 2020, Boeing made a stunning reversal of its previous goal to prevent pilot simulator training and recommended that simulator training be required for all MAX pilots once the plane is ungrounded and returns to service." (p. 141), essentially acknowledging the critical need for such training to reduce the likelihood of future accidents.

Although NASA has always acknowledged the critical need for simulator training for spaceflight missions, research is needed to develop guidelines and standards for the necessary level of simulator and simulation fidelity.

Ground-Based Evidence: Training for Expertise with Automation

In a study of aviation accidents that occurred over a 30-year interval, Strauch (2017) addressed the need for training for automation. Strauch found that, "Automated systems continue to provide capabilities exceeding operator's need for effective system operation and provide interfaces that can hinder, rather than enhance, operator automation-related situation awareness." (p. 204). Strauch investigated 3 separate aviation accidents (including the July 6, 2013, crash of Asiana Flight 214 described later in this report) and argued that, "By not addressing expected operator expertise levels when designing automated systems" aviation accidents have been allowed to continue (p. 205).

Strauch suggested metrics to address the necessary expertise, stating that operators, "must have sufficient expertise to operate systems effectively through routine and nonroutine phases and diagnose the causes of and effectively respond to unexpected system phases and failures." (p. 222) However, Strauch concluded that, "there are few, if any, standards established for the skills needed for automation system operation at the expert of 'qualified' level. The discrepancy between the expertise levels needed to operate systems through the automation and the automation capabilities the system provide also has not been established. Researchers should, hopefully, identify automation capabilities needed for effective system operation but that remain within identified expertise levels." (p. 223-223).

As this evidence shows, research is needed to determine the skill level needed for expertise with automated and intelligent systems, and research is needed to develop guidelines and standards for the level of simulator and simulations fidelity necessary to train for such expertise.

Training Program Design

Ground-Based Evidence: Analysis: Duty and Task Allocation

A job analysis normally results in defining a job based on a set of duties and tasks that one person would reasonably be able to perform in the work environment. Although a job analysis can be difficult to perform, especially for more complex tasks, the process is well understood (Salas et al., 2006). The challenge for NASA is that the number of duties and tasks assigned to each astronaut is more than would normally be allocated to one individual. NASA has identified this as one of the factors contributing to training retention issues (see e.g., Dempsey et al., 2018; Sonoda, 2016); yet, unlike in terrestrial environments, NASA cannot simply increase the workforce and distribute the duties and task among a larger crew. Hence, research is needed to develop a systematic method for allocating duties and tasks among the crewmembers of a future LDEM that lowers the training demand.

Ground-Based Evidence: Design: The Training Flow Across the Training Continuum

As early as 1953, in the early days of human factors research, training had already been recognized as a critical issue. In his seminal work on human factors in air transportation, McFarland (1953) dedicates a whole chapter to training issues. He notes that “little attention has been given to the human factors involving an airman’s understanding of his environment and the physical factors influencing his efficiency” (p. 152). More recent works continue to echo the same concerns. Training deficiencies are clearly indicated in accidents and incidents in aviation (e.g., Barshi & Loukopoulos, 2012; Dismukes et al., 2007; Loukopoulos et al., 2009), other modes of transportation (NTSB), and in other high-risk industries (e.g., Grote, 2009; Helmreich, & Merrit, 1998; Reason, 1997).

Human error has been implicated as a causal factor in nearly 2 thirds of NASA spaceflight mishaps (Chandler, 2007), and similar situations exist in related domains like commercial and military aviation (70-80% of incidents and accidents involve human error directly, and 100% of accidents and incidents involve human limitations in some way, see, e.g., Maurino et al, 1995). In a significant proportion of incidents involving human error, incorrect procedure execution played a role. Procedure execution errors (both of omission and commission) result from some combination of inadequately designed tasks; inadequately designed procedures or tools; incomplete, inaccurate, or difficult-to-use documentation; fatigue, stress, injury, or illness; insufficient training (including lack of training for unanticipated operations); degradation of trained skills or knowledge; or inadequate understanding of the operational environment. A study of aviation accidents that happened between 1988 and 2006 (Velazquez et al., 2015), determined that training inadequacies were the single largest contributing factor to accidents in which crew error was a causal factor—training inadequacies were found to be a contributing factor in 48% of the accidents.

As this evidence shows, research is needed to provide a program of training across the entire training continuum that supports the necessary level of retention and expertise.

Ground-Based Evidence: Design: Methods
Salas et al. (2006) noted that research on design and development of training indicates that variations of practice schedules, the introduction of difficulties during practice, and the nature of feedback “enhances retention and generalization”. Hoffman et al. (2010) also found that the practice factors used in training, such as massed versus distributed factors and procedures versus closed loop tasks, affect the in-mission performance of the trained tasks.

In a controlled research study conducted for the U.S. Army to analyze training methods, technologies, and measures for training soldiers to operate in high stress, complex environments, training experts designed a 4-day capstone course of training based on well-established principles of learning (Wolf, 2017). The training course included 2 days of classroom instruction interleaved with virtual/gaming mission training followed by 2 days intensive immersive live mission training. Training was designed to increase in complexity throughout the course, and the soldiers who undertook the capstone training course significantly increased acquisition of both knowledge and skills: advanced SA increased by 33%, team effectiveness increased by 26%, tactical combat casualty care increased by 40%, and after action reviews increased by 43%. This study clearly shows the impact that an effective program of training with methods based on empirical principles of learning can have on learner performance.

Research is needed to determine which methods best support acquisition, retention and transfer of spaceflight skills and tasks, and guidelines are needed to ensure these methods are effectively employed.

Ground-Based Evidence: Evaluation Methods
Evaluation measures can be used to determine skill acquisition during or at the completion of training. However, the true measure of concern is retention of trained tasks in the operational environment. In a review of skill retention and multi-day training in various fields (medical, military, marine and offshore safety), Sanli and Carnahan (2018) reported that the level of training and skill level at the end of training affected later retention, i.e., the more proficiency achieved during training, the better the retention later. However, Sanli and Carnahan contend that performance during or at end of practice may not be as good a predictor of on the job performance as performance during retention and transfer tests of learning.

Research is needed to define evaluation measures, including the evaluation criteria, the method of evaluation, and the timing of the evaluation, to ensure retention of mission critical tasks.

Performance Support

Issues related to performance support are documented in ground-based high-risk endeavors, where performance support is provided via other team members or via tools, including procedures, hardware, software, intelligent systems, or virtual assistants. In a study by Schaafstal et al. (2000), Navy weapons engineers' problems troubleshooting were attributed to insufficient training and to technical documents written from an engineering viewpoint instead of a maintenance viewpoint; the latter of these is clearly a performance support issue.

Aviation incident and accident reports have shown that inadequate communication between the ground and crew can cause frustration and can adversely affect performance, especially in the absence of effective performance support tools. Many times, crewmembers have not been able to identify information regarding what the ground could assist with and what tasks could be automated to facilitate crew productivity (Rando et al., 2005). The most efficient and effective teams (e.g., aviation, military, design teams) manage to coordinate their activities with just enough, but not too much communication (Entin & Serfaty, 1999; Orasanu & Fischer, 1992; Patrashkova-Volzdosk et al., 2003). However, a team is at risk in the absence of task-specific procedures and appropriate communication training, especially for distributed teams in which shared context and location cannot facilitate communication. To overcome this risk, task procedures, training, and performance support tools must be designed to communicate essential information while keeping the process cost down, thereby minimizing the added workload required to communicate among team members.

Research is needed to develop guidelines for performance support tools that support in-mission task performance.

Designing for Trainability

Designing for trainability is intended to ensure that the vehicle habitat, display interfaces, system, and task designs support skill acquisition and retention, and that the design is not so overly complex (i.e., essentially untrainable given training constraints) that it impacts operational performance. Unfortunately, commercial aviation accidents continue to show the consequences of failing to design for trainability, including the July 6, 2013, crash of Asiana Flight 214 at the San Francisco airport (Figure 37).

Figure 37. Fire damage to the Fuselage of Asiana Flight 214. (NTSB, 2014b)

The NTSB concluded that "The PF [pilot flying] had an inaccurate understanding of how the Boeing 777 A/P [autopilot] and A/T [autothrottle] systems interact to control airspeed in FLCH SPD [flight level change - speed] mode, what happens when the A/T is overridden and the throttles transition to HOLD in a FLCH SPD descent, and how the A/T automatic engagement feature operates. The PF's faulty mental model of the airplane's automation logic led to his inadvertent deactivation of automatic airspeed control. Both reduced design complexity and improved systems training can help reduce the type of error made by the PF." (NTSB, 2014b, p. xii). Strauch (2017) concurred with the NTSB's conclusions on design complexity and training by stating that by, "not addressing expected operator expertise levels when designing automated systems, … and not considering training constraints in automation system design" aviation errors that resulted in accidents, including Asiana Flight 214, have been allowed to continue (p. 205).

With the anticipated increased automation and complexity in future space vehicles, crewmembers are likely to find themselves in similar situations to that of the Asiana Captain. Yet, as both the NTSB report and Strauch's research shows, future systems that are designed considering human cognitive abilities and limitations and training constraints can improve training and performance, and reduce the likelihood of catastrophic errors. The challenge for NASA is that guidelines for designing for trainability do not yet exist. Research is needed to develop HSI design principles to support trainability of in-mission tasks.

Summary of Ground-Based Evidence for Training Contributing Factors to the HSIA Risk
Unlike evidence from spaceflight, there is an abundance of ground-based evidence based on controlled studies related to training factors that contribute to the HSIA risk. This evidence, along with the ground-based case studies and expert opinion, show that without an effective program of training that supports the retention and transfer necessary for in-mission performance, and the level of expertise needed to operate a complex vehicle and respond to unanticipated critical malfunctions, there is a likelihood of catastrophic consequences.

However, much of the literature on ground-based training is motivated by current training practices aimed at current training needs. Although some experience exists with operations in extreme environments on Earth, no experience exists with long-duration space missions during which the crew must retain skills trained pre-mission over long periods of time, must have the expertise to conduct semi-autonomous to autonomous operations, and must accommodate significant communication delays with ground, and where so little is known about the environment. Thus, not only must known deficiencies in the current training be resolved, robust methods and tools must be developed to prepare crews for the unknown. Although existing ground-based research reveals general challenges that are relevant to long-duration, high-autonomy missions, research is needed to provide future crewmember with the training needed for autonomous and semi-autonomous mission operations.

Risk of Inadequate HSIA and Mission, Process, and Task Design Factors

Mission tasks, task flows, schedules, and procedures must be developed to provide adequate task design in increasingly autonomous operations during LDEMs. We need to understand relevant human capabilities and limitations for performing tasks and how these may impact workload and degrade performance during LDEMs. We also need to understand the effect that other factors may have on human-system performance when dealing with off nominal situations autonomously (e.g., automation, autonomous operations, reality-augmenting devices devices). The HSIA is rooted heavily in the requirements, policies, and design processes of the task environment.

Requirements, policies, and design processes

Defining tasks without an understanding of the task environment leads to inadequate task design. Therefore, analyzing and understanding realistic operations concepts is crucial to providing procedures that work effectively and support safety and efficiency. Loukopoulos et al. (2009) significantly reduced pilot errors by conducting a systematic revision of aircraft cockpit procedures. Procedure revisions were initiated with a task analysis viewpoint, which focused on identifying concurrent tasks although they had been designed as if they were to be accomplished in isolation. The authors focused on multitasking, i.e., handling more than one task at the same time. They analyzed the sequence of tasks performed by the captain and first officer, from preflight through all flight phases, including power-down, and found that procedures for these tasks are typically written for the crew only, omitting the complexities introduced by the interactions between ground support and the flight crew. This research showed that the operating environment for the procedures was linear, predictable, and controllable, whereas the real environment included a variety of disturbances requiring changes to the task sequence. Many of the disturbances were unpredictable or uncontrollable. The researchers also identified many crew omissions. Four distinct patterns frequently resulted in these types of errors: interruptions and distractions, tasks that cannot be executed in the planned sequence, unanticipated new tasks, and multiple tasks that must be interleaved. The authors supported one airline's program to review and revise its procedures, and reduced error rates by over 80%. In addition, the revised procedures were learned more easily, improving training efficiency. This human error research, within the context of real-world operations, illustrates the importance of designing procedures that consider the complete system.

Time between the training and the use of the knowledge may also contribute to performance. Inadequate crew knowledge of automated systems was a factor in more than 40% of commercial aviation accidents and 30% of serious incidents between 2001 and 2009 (Abbott, 2010, oral presentation summarized in Rosenkrans, 2010). Abbott claimed that flight crews are not properly trained for modern cockpits and that a radical change in their recurrent training and in the standard operating procedures of airlines were needed. Although this risk is outlined in the context of commercial aviation, it is highly relevant for LDEMs.

Human Capabilities and Limitations

Workload

On American Airlines Flight 965, an automation failure during a critical phase of flight increased the pilot workload and, ultimately, caused a crash killing 159 passengers (Ladkin, 1996). In 1976, a mid-air collision between British Airways Flight 476 and Inex-Adria Aviopromet Flight 550 killed all passengers and crew on both planes. The air traffic controller responsible for directing the planes in a congested airspace was overloaded with information. The controller was working alone without an assistant controller. Upon realizing the collision course (which would have actually been a near-miss), the air traffic controller started talking in his native Croatian language, likely due to extreme stress, and did not realize the British Airways pilot could not understand. The controller accidently directed the planes into a collision course (AAIB, 1977).

Situation Awareness (SA)

As stated by Jones and Endsley (1996), SA-related human errors, which may be due in part to misperception or misinterpretation of information, are a significant factor in many aviation incidents. Indeed, over 143 aviation incidents between January 1986 through May 1992 were found to be related to SA errors, often with multiple errors occurring per incident (262 total SA errors were found in 143 incidents). SA errors can be classified as level 1 SA errors (failure to perceive or misperception), level 2 errors (improper integration or comprehension of information, i.e. misinterpretation), and level 3 errors (incorrect projection of future actions). Of the 262 SA errors evaluated, 8.7% were explicitly associated with level 1 errors of misperception, whereas 21.1% of pilot SA errors and 17.2% of air traffic controller SA errors were level 2 errors associated with misinterpretation of information (Jones & Endsley, 1996).

Operational Tempo

Operational tempo is often driven by the pressure to stay on schedule irrespective of the consequences of an inappropriate decision. For instance, the Koninklijke Luchtvaart Maatschappij (KLM)/Pan Am collision at Tenerife Airport, Spain, in 1977 occurred when the pilot of one of the planes took off without having received clearance (Air Line Pilots Association [ALPA], 1977). Evidence suggests that this may have been caused by time constraints imposed by deteriorating weather, ambiguous communication between the flight crew and control tower, non-standard operating procedures, and role and responsibility issues on the flight deck.

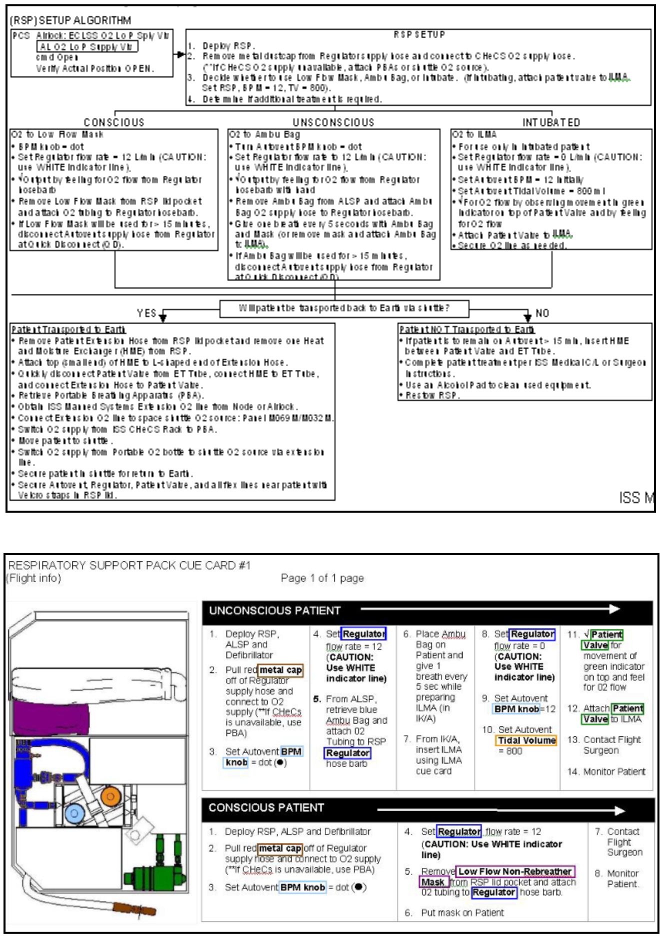
Operational tempo may also be affected by the ability of the person to complete a task uninterrupted. Common interruptions in air traffic control clearance during runway approaches may lead to increases in pilot procedure performance errors, including omission of tasks, and increases in procedure completion times (Latorella, 1996).

Procedural Guidance

Desaulniers et al. (1988) investigated the impact of various display variables on the effectiveness of computer-based procedure displays. Procedures were presented in text, extended text, and flowchart formats. Text and extended text were structured prose formats differing in the spatial density of presentation. The flowchart format differed from the text format in both syntax and spatial representation. Subjects were asked to use the procedures to diagnose a hypothetical system anomaly, and accuracy was higher with the flowchart format. Although overall task completion times did not differ across formats, the flowchart format required significantly less time to implement steps. Furthermore, the follow-on study showed that completion times for flowchart procedures decreased with increasing window size; however, accuracy of performance decreased substantially. These studies show that the presentation of procedures strongly affects performance.

Many airplane incidents illustrate the importance of good procedure design. In 1987, Northwest Airlines Flight 255 (NTSB, 1987b) crashed shortly after takeoff, killing all but one passenger. As the plane lifted off the runway, it began to roll from side to side, stalled, rolled into a pole at the end of the runway, and then rolled onto a road, hitting vehicles and bursting into flames when it hit an overpass. The NTSB concluded that the probable cause of the accident was the failure of the pilots to use the taxi checklist to ensure the aircraft flaps were properly configured for takeoff.

Performance deficit due to inadequate ISS task design was illustrated during a ground-based study to test the usability of the procedure (as written on a “cue card,” Figure 38) for the Respiratory Support Pack (RSP) (Hudy et al., 2005). The RSP was designed for use during medical contingencies involving respiratory distress, and it was therefore expected that the complicated RSP cue card procedure would be used in time-critical situations when a crewmember’s life could depend on the outcome. Data collected as subjects executed the procedure checklists demonstrated that some procedures, equipment design, and labeling could induce errors and ultimately risk crew health. The procedures and the sequence of equipment use did not enable a crewmember to establish a patient's airway in the time necessary to prevent irreversible brain damage. The Crew Medical Officer typically receives very limited training on use of this equipment, and the cue cards thus hold vital information on how to execute the procedures. This example illustrates the importance of appropriate procedures and support for training to ensure that tasks can successfully be performed, especially in the case of an emergency. The cue card was subsequently redesigned to support the task (Figure 37).



(RSP) SETUP ALGORITHM

PCS Airlock: ECLSS O2 Lo P Sply Vlv
AL O2 Lo P Supply Vlv
cmd Open
Verify Actual Position OPEN.

RSP SETUP
1. Deploy RSP.
2. Remove metal dustcap from Regulator supply hose and connect to CHeCS O2 supply hose. (**If CHeCS O2 supply unavailable, attach PBAs or shuttle O2 source).
3. Decide whether to use Low Flow Mask, Ambu Bag, or Intubate. (If intubating, attach patient valve to ILMA. Set RSP, BPM = 12, TV = 800).
4. Determine if additional treatment is required.

CONSCIOUS

O2 to Low Flow Mask
- BPM knob = dot
- Set Regulator flow rate = 12 L/min (CAUTION: use WHITE indicator line).
- √Output by feeling for O2 flow from Regulator hosebarb
- Remove Low Flow Mask from RSP lid pocket and attach O2 tubing to Regulator hosebarb.
- If Low Flow Mask will be used for > 15 minutes, disconnect Autovent supply hose from Regulator at Quick Disconnect (QD).

UNSCONSCIOUS

O2 to Ambu Bag
- Turn Autovent BPM knob = dot
- Set Regulator flow rate to 12 L/min (CAUTION: use WHITE indicator line).
- √Output by feeling for O2 flow from Regulator hosebarb with hand
- Remove Ambu Bag from ALSP and attach Ambu Bag O2 supply hose to Regulator hosebarb.
- Give one breath every 5 seconds with Ambu Bag and Mask (or remove mask and attach Ambu Bag to ILMA).
- If Ambu Bag will be used for > 15 minutes, disconnect Autovent supply hose from Regulator at Quick Disconnect (QD).

INTUBATED

O2 to ILMA
- For use only in intubated patient
- Set Regulator flow rate = 0 L/min (CAUTION: use WHITE indicator line).
- Set Autovent BPM = 12 initially
- Set Autovent Tidal Volume = 800 ml
- √For O2 flow by observing movement in green indicator on top of Patient Valve and by feeling for O2 flow
- Attach Patient Valve to ILMA.
- Secure O2 line as needed.

Will patient be transported back to Earth via shuttle?

YES

Patient Transported to Earth
- Remove Patient Extension Hose from RSP lid pocket and remove one Heat and Moisture Exchanger (HME) from RSP.
- Attach top (small end) of HME to L-shaped end of Extension Hose.
- Quickly disconnect Patient Valve from ET Tube, connect HME to ET Tube, and connect Extension Hose to Patient Valve.
- Retrieve Portable Breathing Apparatus (PBA).
- Obtain ISS Manned Systems Extension O2 line from Node or Airlock.
- Connect Extension O2 line to space shuttle O2 source: Panel M069M/M032M.
- Switch O2 supply from ISS CHeCS Rack to PBA.
- Move patient to shuttle.
- Switch O2 supply from Portable O2 bottle to shuttle O2 source via extension line.
- Secure patient in shuttle for return to Earth.
- Secure Autovent, Regulator, Patient Valve, and all flex lines near patient with Velcro straps in RSP lid.

NO

Patient NOT Transported to Earth
- If patient is to remain on Autovent > 15 min, insert HME between Patient Valve and ET Tube.
- Complete patient treatment per ISS Medical C/L or Surgeon instructions.
- Use an Alcohol Pad to clean used equipment.
- Restow RSP.

ISS M

RESPIRATORY SUPPORT PACK CUE CARD #1
(Flight info)
Page 1 of 1 page

UNCONSCIOUS PATIENT

1. Deploy RSP, ALSP and Defibrillator
2. Pull red **metal cap** off of Regulator supply hose and connect to O2 supply (**If CHeCs is unavailable, use PBA)
3. Set Autovent **BPM knob** = dot (●)
4. Set **Regulator** flow rate = 12 (**CAUTION: Use WHITE indicator line)**
5. From ALSP, retrieve blue Ambu Bag and attach 02 Tubing to RSP **Regulator** hose barb
6. Place Ambu Bag on Patient and give 1 breath every 5 sec while preparing ILMA (in IK/A)
7. From IK/A, insert ILMA using ILMA cue card
8. Set **Regulator** flow rate = 0 (**CAUTION: Use WHITE indicator line)**
9. Set Autovent **BPM knob**=12
10. Set Autovent **Tidal Volume** = 800
11. √ **Patient Valve** for movement of green indicator on top and feel for 02 flow
12. Attach **Patient Valve** to ILMA
13. Contact Flight Surgeon
14. Monitor Patient

CONSCIOUS PATIENT

1. Deploy RSP, ALSP and Defibrillator
2. Pull red **metal cap** off of Regulator supply hose and connect to O2 supply (**If CHeCs is unavailable, use PBA)
3. Set Autovent **BPM knob** = dot (●)
4. Set **Regulator** flow rate = 12 (**CAUTION: Use WHITE indicator line)**
5. Remove **Low Flow Non-Rebreather Mask** from RSP lid pocket and attach 02 tubing to **Regulator** hose barb.
6. Put mask on Patient
7. Contact Flight Surgeon
8. Monitor Patient.

Figure 38. RSP Information Card Before Evaluation and After Evaluation and Redesign

Issues related to usability and human capabilities can also arise when safety precautions are factored into design for on-orbit operations, but not for preflight ground activities. On January 27, 1967, when the Apollo 1 crew initiated what should have been a nominal countdown drill, disaster struck and fire erupted in the command module at Kennedy Space Center's Pad 34. All 3 crewmembers lost their lives because they were unable to open the hatch and escape the command module to safety (Kranz, 2000). Many factors contributed to this incident, including the inadequate hatch design. The hatch was designed to open inward, which was impossible for a human to open at the pressure levels within the vehicle while it was on the ground. Procedures or processes were not in place to deal with this type of emergency event. Therefore, crewmembers did not have the necessary procedural or technical guidance to handle an event of this magnitude, which led to the worst possible consequence: loss of life. Little consideration had been given to risks and hazards associated with any of the preflight activities, only those pertaining to spaceflight,

The way information is presented can greatly affect the retention of procedural and technical data and determine operations success or failure. A study collected feasibility data for 2 prototype display types for viewing real-time medical procedures and just-in-time (JIT) training. A hand-held personal digital assistant and an HMD system were compared while performing a simulated JIT medical procedure using ISS flight-like equipment (Figure 39). Participant reported that both devices provided too little information to aid SA of how far into the procedure they were.

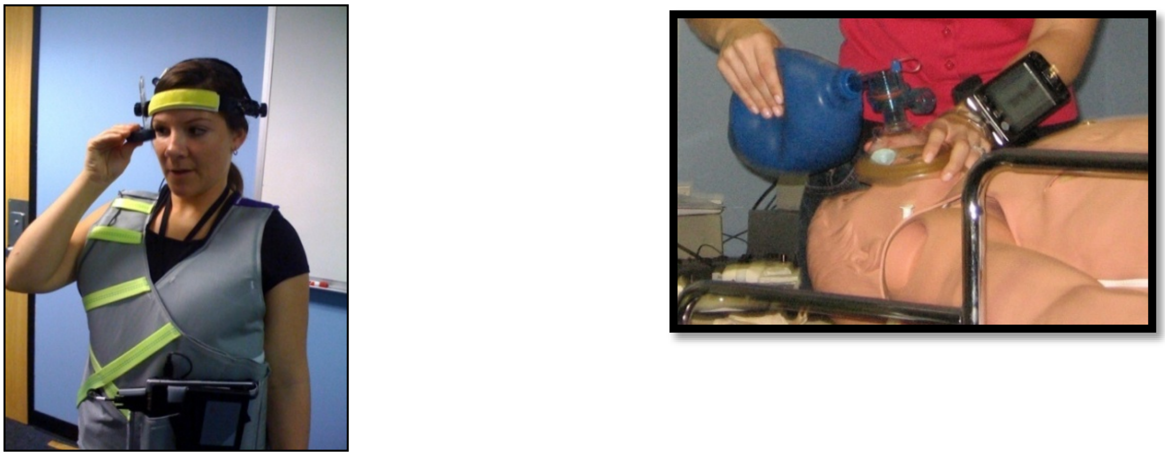


Figure 39. Participants Perform Simulated JIT Medical Procedures

Multimodal enhanced electronic procedures allow for powerful new methods of user-procedure interaction that have yet to be fully leveraged. Enhanced procedures can include the use of augmented or VR devices, which increase user immersion and interaction capabilities with procedures. Although promising, the utility of these devices to improve SA and workload remains unproven. Research is ongoing at NASA to compare performance metrics during the use of AR procedure devices and during the use of more traditional electronic procedure devices (such as a tablet device). Additional work is needed to adequately identify the implications of using such devices for electronic procedures.

Another study assessed multiple versions of checklist information by testing for differences in performance and user opinion of the current NASA *ISS Integrated Medical Group (IMG) Medical Checklist* (JSC-48522), and of 2 alternatively organized versions. Each version contained the organization of the emergency and non-emergency medical procedures and diagnoses. The original checklist version featured *both* alphabetical and anatomical organization of the medical procedures. The 2 alternate versions offered an anatomical organization and an alphabetical organization. The results showed that the anatomical and alphabetical versions led to shorter task completion times than the checklist version. More recently a new cognitive aid device and a general toolset for procedure aids have been developed, along with a taxonomy of cognitive aid design principles considered to be most influential to user task performance.

Training for tasks that will be performed under extreme conditions and environments can lead to less than adequate procedural knowledge. This can be because simulated environments and ground-based, full-scale models or mockups cannot be completely representative of flight conditions. Representing a true 0g environment on the ground has presented many challenges for training and information presentation; thus, simulations may provide inadequate information during preflight training.

HSIA and Exploration Class Mission Task Decomposition

An important report by Stuster et al., (2018) describes many of the issues embedded within HSIA. Before beginning the design and development of any major space program, the tasks crewmembers will need to perform to accomplish the mission must be determined. This activity is called Task Analysis and is part of NASA's HCD process (NASA, 2014). This process includes identifying tasks the crew must perform and important task-related information including crew roles and responsibilities; task dependencies, duration, and frequency; necessary clothing and equipment; constraints; and knowledge, skills, abilities, and training required to perform the task. To assess expected tasks and abilities on a LDEM, Stuster et al., (2018) used the critical abilities and tasks method to create an exhaustive list of likely tasks to be performed during the mission. A total of 1,125 tasks were created in this process, and subsequently summarized into 158 representative summary task statements to decrease the time required by SMEs to rate tasks. The table of all 1,125 tasks and 158 summary task statements about the likely tasks during LDEMs, skills and abilities needed therein can be found in the appendix of Stuster et al.'s 2019 publication. These task functions were then used to inform the development of 8 primary crew specialties: Leader, Biologist, Geologist, Physician, Electrician, Pilot/Navigator, Mechanic/Engineer, and Computer Specialist.

A total of 72 SMEs, representing the technical specialties identified during the task analysis, performed the ability modified card-sort for their respective roles; 6 SMEs ranked the 121 abilities for more than one technical specialty and 42 of the SMEs also ranked the abilities for the role of expedition leader. The 8 crew specialties/roles include leader, biologist, geologist, physician, electrician, pilot/navigator, mechanic/engineer, and computer specialist. The analyses indicated that the work encompassed by the 1,125-tasks identified during the study could be performed by the 8 primary crew specialties/roles, with 4 ancillary roles. The report identified a 4-, 5-, and 6-person hypothetical crew composition, using a rational process based on results of the task analysis and ability rankings to illustrate a potential assembly of crew to satisfy operational requirements.

A 3-year expedition to Mars will test the endurance of both humans and their technology. Approximately 10% of the 1,125 tasks the crew of a Mars expedition must be prepared to perform involve maintenance and repair, from the spacecraft to the software that will control all systems. Ensuring the reliability of technological components will reduce maintenance workloads and overall risk, and will also increase the crew's confidence in systems and the probability of mission success. Engineers use 4 basic design strategies for increasing reliability and reducing the risk and effects of component or system failure: redundancy, overbuilding, graceful degradation, and maintainability. All 4 of these strategies should be employed to increase the reliability of mechanical, electrical, and software components of LDEMs.

Additional implications of the task analysis (and results of other research) on crew-selection and training are: (1) Provide *training* in Mars geology to all crewmembers (Eppler et al., 2016); (2) Ensure that at least 2 crewmembers develop *proficiency* operating the prepositioned excavation and well-drilling equipment; (3) The pilot and backup should lead robot and rover operations, but other crewmembers also should be *trained*; (4) *Train* a crewmember to lead the observation of the Transit of Earth and Moon during the Cruise to Mars, but conduct the research as a group activity to foster solidarity early in the expedition; and, (5) Require that all crewmembers and mission control personnel receive training concerning the behavioral effects of isolation and confinement and the need to monitor each other's adjustment to the conditions.

Requirements for Successful Autonomous Missions

As a reminder, in Phase I of the HCAAM project (Holden et al., 2019b), the investigators performed a gap analysis of NASA-STD-3001 and a standards/guideline literature search to arrive at a consensus review on human system standards and guidelines documents from key agencies, including the DOD, the FAA, the DOT, and the Nuclear Regulatory Commission. Candidate standards and guidelines were selected to fill gap areas in the NASA document set. In Phase II, the focus turned to crew tasks. Investigators reviewed documents such as *Generalizable Skills and Knowledge for Exploration Missions* (Stuster et al., 2018; Stuster et al., 2019), and *Gateway Concept of Operations* (Urbina & Fleming, 2019) to develop a high-level draft list of tasks that crew might perform on a mission to Mars, focusing on tasks that might prove challenging for an autonomous crew. Once the list of tasks was drafted, an expert focus group meeting was held to validate and gain consensus on the task list. Participants included representatives from crew, flight/mission control, mission planning, medicine, behavioral health, and training. Facilitators assessed the draft list of autonomous tasks and participants provided valuable feedback that validated and expanded the draft autonomous task list. Participants also shared their primary concerns for LDEMs, which revealed certain trends, including: (1) a lack of preparation and experience for autonomous crew operation, (2) insufficient intelligent system testing, and (3) inadequate training approaches for LDEMs.

8 INTERDISCIPLINARY FACTORS - CROSS-RISK DEPENDENCIES AND EVIDENCE

The risk of inadequate HSIA encompasses all the former risks from the Human Factors Portfolio together into a conceptual framework proposed to address the integration of onboard capability and crew roles and responsibilities necessary to enable a crew to respond effectively and efficiently in the required increasing autonomous mission operations framework. Factors such as sleep loss, work overload, cognitive impairment due to medical conditions, social isolation, operational/task related stressors, and communication also relate to the HSIA risk because they affect crewmembers' abilities to communicate, share information, and interact effectively with computer-based systems. The risk of Impaired Manual Control can also contribute to the HSIA risk. Microgravity, vibration, and deconditioning can affect crewmembers' ability to perform fine motor control tasks. Impairment of fine motor control will affect crewmembers' abilities to interact with input devices, such as cursor control devices. The HSIA risk also interacts with risks from Behavioral Performance; the Risk of Adverse Behavioral Conditions and Psychiatric Disorders includes contributing factors of sleep loss, work overload, cognitive impairment due to medical conditions, and operational/task-related stressors. These factors impact crew workload and task performance and could contribute to the HSIA risk.

Dependencies and Interrelationships

The ExMC Element is performing work related to HSIA. In a report of operational concepts developed for a medical system recommended for the Gateway Missions (NASA, 2019a), researchers convey that "a well-designed Gateway Habitat Medical System lowers mission medical risk, in part, by minimizing training time and operational complexity" (p. 12). A concept of operations has been developed by ExMC researchers that provides a vision of medical care for Mars exploration missions (Urbina & Fleming, 2019) in which a physician astronaut will serve as a director of care during real-time medical events. Resources will be required in the form of onboard medical references, decision support systems, and task training.

Fundamentally, spaceflight operations will depend on astronauts retaining preflight cognitive performance levels during long-duration spaceflight. If cognitive performance decreases during spaceflight, human performance using automation and/or robotics is affected. Space-specific stressors can affect cognitive performance (microgravity, dark-light cycle, and radiation exposure), as can space-relevant stressors (including isolation, noise, air quality) (Sandal, 2018). Currently, the limited spaceflight evidence (in LEO) does not support the existence of generalized cognitive deficits during spaceflight (Casler & Cook, 1999; Strangman et al., 2014). However, this is contrary to the anecdotal reports from spaceflight (Schroeder & Tuttle, 1992; Slack et al., 2016). Furthermore, radiation's detrimental effects on the CNS are of concern for missions beyond LEO. The cognitive processes that govern goal-directed action and adaptive responses to novel, complex, or ambiguous situations are more likely to affect human-automation and human-robotic interaction.

The risk of inadequate HSIA also encompasses the effects of stressors such as sleep deprivation and fatigue; fatigue that, in turn, may lead to performance errors and high workload that could compromise mission objectives and, consequently, the mission itself. Research to mitigate this risk may include developing a tool to self-assess cognitive function and fatigue; light therapy to improve phase shifting, alertness, and mood disorders; individualized protocols for sleep-wake medication use; sleep dose-response recovery curves and individualized models to implement CMs and optimal work-rest schedules; and other evidence-based means to improve individual sleep quality and reduce fatigue. A recent meta-analysis indicates that sleep disruption affects performance of complex tasks less than it does performance of simple task (Wickens et al., 2015b).

Not enough research has directly investigated the relationship between sleep function and performance of human-automation or human-robotic systems. Research is ongoing to validate ROBoT-r task performance as a measure for behavioral performance under analogous spaceflight settings (including isolation, stress and fatigue conditions) but those results are yet to be published. Researchers have assessed how fatigue affects other components of cognition, particularly executive cognition. Tucker and colleagues challenge the idea that sleep deprivation primarily affects cognitive processes that rely on attention (Tucker et al., 2010). Fatigue affects cognitive flexibility distinctly from impairments in vigilant attention (Honn et al., 2018; Whitney et al., 2014). Feedback blunting describes the process that results when understanding of circumstances is not updated based on new information, which affects decision-making and induces performance deficits. Failure to consider new information due to fatigue could lead to accidents and mishaps, and is particularly dangerous in conjunction with automation bias and overreliance. Future NASA missions will expect users of automated and robotic systems to be responsible for high-level, strategic supervision and management. We know that these capabilities are affected by sleep deprivation, however, little research has been conducted on conditions directly relevant to future LDEMs.

9 COMPUTER-BASED MODELING AND SIMULATION

Computer-based modeling and simulation can be used early in the design process to complement user-based research. Formal methods of modeling and evaluation use existing data about physical and cognitive activities to estimate human task completion times and errors, among other performance measures. Modeling and simulation can be used to understand specific tasks and workload levied on human operators; however, these kinds of modeling and simulation methods are not readily available. In the human factors engineering domain, modeling and HITL evaluations must be used in concert. No high-fidelity human performance models (HPMs) exist, and most of the existing models have not been sufficiently validated or certified. Accordingly, models must be used in a limited fashion, i.e., to help determine the critical areas that should be addressed through the more costly, but more representative, HITL evaluations. HPMs, when used in concert with HITL evaluations, can be quite effective and cost efficient in habitat design, but accurately modeling the human is extremely difficult, especially when considering the addition of spaceflight-specific complexities such as varying gravity environments and bulky EVA suits. A need exists for higher-fidelity human performance models, validation of existing models, and better integration of models that portray various aspects of human performance.

NASA has performed a wide variety of human factors-related modeling and simulation work, including human anthropometric modeling, habitat volumetric analysis and layout design (SOLV; Thaxton et al., 2017; Chen et al., 2020), spacecraft interior lighting and acoustic modeling, and human task performance modeling (MIDAS, Gore, 2010; MIDAS-FAST, Gacy et al., 2011; Sebok et al., 2013; S-PRINT, Wickens et al., 2014a,b,; 2015a,b,c,d).

10 RISK IN CONTEXT OF EXPLORATION MISSION OPERATIONAL SCENARIOS

Future exploration mission scenarios will increase in duration and in distance from Earth. This will require new technology, new work methods, and new ways of ensuring that these novel elements are suitably integrated. LDEMs will require greater flexibility and less dependence on ground support, and new methods of interaction between ground-based resources and the crew. Designing spacecrafts to be operated beyond LEO or on the lunar surface, where communications may be interrupted by malfunctions, requires careful design, particularly for the tasks needed to maintain the spacecraft and complete the mission objectives. Knowledge can be provided from other sources, such as onboard databases or mission support on Earth. However, tasks cannot require more immediate knowledge than can be accessed by 4 individuals. Careful use of automation will be required to augment the capabilities of this small crew. Routine tasks that consume a great deal of time can be automated to allow crewmembers to perform the tasks that require uniquely human judgment, creativity, and problem-solving skills. It is critical that performance operating limits (POLs), and the permissible exposure limits (PELs) are correctly specified and the HFBP Element is poised to provide these POLs and PELs for the crew for these LDEMs.

Automated planning capabilities and tools are required for increased autonomy during lunar and Mars missions. These tools could provide necessary automated support to determine alternative plans and solutions for managing daily tasks. Current spaceflight crews rely on onboard automated systems, and crews of future missions with increased flight duration and increased autonomy will rely even more on these systems to provide information that is appropriate, accurate, and up to date. Increased automation will result in the need for additional training to ensure that the crew can perform the automated tasks in the event of automation failure. Automated tasks must be carefully designed to prevent the crew from becoming unaware of, or complacent about, potential hazards—situation could ultimately result in system errors, degraded crew performance, and compromised crew and vehicle safety.

11 GAPS

Eight major gaps have been identified that relate to the risk of inadequate HSIA:

1. HSI crew health and performance outcomes, measures, and metrics;
2. Future exploration habitat/vehicle systems and mission scenario risk factors to crew health and performance;
3. Onboard systems to maintain and monitor crew health and performance, and thresholds to determine countermeasure (CM) implementation;
4. HSI in vehicle/habitat and computer interface Design Phase, to mitigate potential performance decrements, for exploration;
5. HSI in development of dynamic and adaptive mission procedures and processes for exploration missions;
6. HSI in development of preflight and onboard training regimens to reduce demands (e.g. neurocognitive, time) on crew, support meaningful work, and mitigate potential individual & team decrements during exploration;
7. Human-automation-robotic systems integration to monitor and enhance crew health & performance, for exploration missions;
8. Synergistic, integrative risk relative to HSIA and onboard CM implementation.

To close the HSIA gaps, operational onboard data collection on workload, usability, task design, and performance is required. This will help to further characterize the risks and inform development of mitigations so that we can be better prepared to support crew on future LDEMs.

The following are research issues critical to supporting the HSIA, research issues that will help bridge the knowledge gaps.

- Conduct research to characterize the effects of spaceflight environmental conditions on higher-level cognitive functions. The existing data are sparse, and they do not clarify the sources of influence. For example, some argue that when a pilot's is spatially disoriented, her/his instinct to regain spatial orientation can drain cognitive resources and that may be the cause of cognitive decrement under disorientation (Gresty et al., 2008).
- Conduct research to determine crew tasks to be performed and skills required for the dynamic phases, including docking and landing. Receiving real-time assistance from the ground will not always be possible for the Artemis crew, especially because communication between Earth and the South Pole of the Moon may have long periods of blackouts (Vanoutryve et al., 2010). To be able to accurately assess crew operational performance in relation to the HLS, research is needed to determine crew tasks to be performed and skills required (e.g., Stuster et al., 2019).
- Conduct research to determine whether cognitive batteries predict operational performance. Fitness to perform is typically assessed using cognition test batteries focused on component cognitive abilities and motor control skills (e.g., Beard & Ahumada, 2019). It is not clear whether and how performance of component skills

predicts performance of higher cognitive functions, and whether and how the latter ultimately predict operational performance.

12 CONCLUSION

New, increasingly autonomous missions will require an evolution of our current spaceflight products and practices. The safety and efficiency of the crew throughout a LDEM will depend to a large extent on their ability to acquire, process, and make adequate decisions using primarily the computer-based information they are provided with while living and working on-orbit. For this to occur, the crew must be presented with the appropriate format and quantity of information needed to allow them to successfully complete assigned tasks. In addition, the information must be provided in a clear, concise, and timely manner, and the method of interaction with the information must be compatible with human capabilities and limitations. Mission planners, procedure writers, and hardware/software designers must be cognizant of this when designing systems that will interact with a human operator.

This evidence report has focused on the factors contributing to a risk of inadequate HSIA. If these contributing factors can be better characterized through research, and if there is an awareness of these factors and mitigations in place, the risk of human error will be reduced significantly. The evidence includes examples from spaceflight, aviation, maritime, and ground-based research that illustrate the effects of several contributing factors to the risk of inadequate HSIA. In some cases, incidents resulted when contributing factors were not considered in design. In other cases, research indicates an awareness of the factors that contribute to the risk, and attempts are being made to prevent or mitigate potential errors. Often a combination of contributing factors is at work—a better understanding of the combined effects to required to develop effective mitigations. Further research is needed to develop transparent methods of measuring HSIA for LDEMs, and tools to improve information presentation, acquisition, processing, and decision making in the increasingly autonomous environment. Without these improvements, errors due to inadequate HSIA will continue to pose a risk to mission success.

13 REFERENCES

Air Accidents Investigation Branch (AAIB) (1992). *Aircraft accident report 4/92 - Report on the incident to British Aerospace ATP, G-BMYK, 10 miles north of Cowly near Oxford, on 11 August 1991*. https://assets.publishing.service.gov.uk/ media/ 542301a2ed915d1371000b4b/4-1992_G-BMYK.pdf

Abbott, K. (2010, November). *Flight path management.* [Oral presentation] 2010 Flight Safety Foundation International Aviation Safety Seminar, Milan, Italy.

Adams, J. A., Deloach, S. A., & Scheutz, M. (2014). Shared mental models for human-robot teams. *Proceedings of the AAAI Fall Symposium*. Technical Report, FS-14–01, pp. 5-6.

Adelstein, B., Beutter, B., Kaiser, M., McCann, R., & Stone, L. (2009a) Effects of transverse seat vibration on near-viewing readability of alphanumeric symbology. NASA technical memorandum, NASA/TM 2009-215385.

Adelstein, B., Beutter, B., Kaiser, M., McCann, R., Stone, L., Anderson, M., Renema, F., & Paloski, W. (2009b). Influence of combined whole-body vibration plus G-loading on visual performance. NASA/TM 2009-215386. https://ntrs.nasa.gov/ citations/20190001113

Adelstein, B., Beutter, B., Kaiser, M., McCann, R., Stone, L., Holden, K., & Root, P. (2009c). Determining subjective thresholds for display usability under combined whole-body vibration plus G-loading. https://humanresearchroadmap.nasa.gov/tasks/task.aspx?i=622

Adelstein BD, Kaiser MK, Beutter BR, McCann RS, Anderson MR (2013) Display strobing: an effective countermeasure against visual blur from whole-body vibration. Acta Astronautica 92:53-64

Ainsworth, L. K. (2004). Task analysis. In C. Sandom & R. Harvey (Eds.), *Human factors for engineers*, pp. 81–112. London: Institution of Engineering and Technology.

Aiple, M., & Schiele, A. (2013). Pushing the limits of the CyberGrasp^{TM} for haptic rendering. *Proceedings of the 2013 IEEE international conference on robotics and automation*. Karlsruhe, 3541-3546. DOI: 10.1109/ICRA.2013.6631073.

Air Line Pilots Association; ALPA (1977). Human factors report on the Tenerife accident. http://project-tenerife.com/engels/PDF/alpa.pdf

Alderson, D. L., & Doyle, J. C. (2010). Contrasting views of complexity and their implications for network-centric infrastructures. *IEEE transactions on systems, man, and cybernetics - Part A: Systems and humans*, 40(4), 839–852. https://doi.org/10.1109/ TSMCA.2010.2048027

Aleksandrovskiy, Y. A. (1976). *States of psychic deadaptation and their compensation*. Moscow: Nauka Press.

Altman, I. (1979). *Privacy as an interpersonal boundary process*. In M., von Cranach, K., Foppa, W., Lepenies & D., Plogg (Eds), *Human ethology*. Cambridge University Press.

American Conference of Governmental Industrial Hygienists (1993) *Threshold limit values for chemical substances and physical agents and biological exposure indices.* American conference of governmental industrial hygienists, Cincinnati, OH.

Anandan, T. (2013, June 6). Robotics industry insights—The end of separation: Man and robot as collaborative coworkers on the factory floor. http://www.robotics.org/content-detail.cfm/Industrial-Robotics-Industry-Insights/The-End-of-Separation-Man-and-Robot-as-Collaborative-Coworkers-on-the-Factory-Floor/content_id/4140

Antonsen, E., Bayuse, T., Blue, R., Boley, L., Hailey, M., Kerstman, E., Mulcahy, R., Nusbaum, D., & Reyes, D. (2018). *Development of an accepted medical condition list for exploration medical capability scoping.* Exploration medical capability (ExMC): working Group participants. [Unpublished report].NASA's Human Research Program.

Antonsen, E., Bayuse, T., Blue, R., Daniels, V., Hailey, M., Hussey, S., Kerstman, E., Krihak, M., Latorella, K., Mindock, J., Myers, J., Mulcahy, R., Red, R., Reyes, D., Urbina, M., & Alston, M. (2017). *Evidence report: Risk of adverse health outcomes and decrements in performance due to in-flight medical conditions.* NASA. https://humanresearchroadmap.nasa.gov/Evidence/reports/Medical.pdf

Anvari, M., Broderick, T., Stein, H., Chapman, T., Ghodoussi, M., Birch, D. W., Mckinley C., Trudeau, P., Dutta, S., & Goldsmith, C. H. (2005). The impact of latency on surgical precision and task completion during robotic-assisted remote telepresence surgery. *Computer aided surgery*, *10*(2), 93-99. https://doi.org/10.3109/10929080500228654.

Ashcraft, N., & Scheflen, A. E. (1976). *People space: The making and breaking of human boundaries*. Anchor, Doubleday, Garden City, N.Y.

Aviation Week and Space Technology. (1995). *English translation of the French Commission of Investigation preliminary report on the June 30, 1994 Airbus A330 Accident, In six parts*. New York, NY: McGraw-Hill Companies.

Aziz, S. (2010). Lessons learned from the STS-120/ISS 10A robotics operations. *Acta Astronautica*, *66*(1-2), 157-165. https://doi.org/10.1016/j.actaastro.2009.05.024

Aziz, S. (2013). Development and verification of ground-based tele-robotics operations concept for Dextre. *Acta astronautica*, *86*, 1-9. https://doi.org/10.1016/j.actaastro.2011.11.004

Baddeley, A. (1992). Working memory. *Science (New York, N.Y.)*, *255*(5044), 556-559. https://doi.org/10.1126/science.1736359

Baggerman, S. D., Rando, C., & Duvall, L. E. (2004). Habitability and human factors: lessons learned in long duration spaceflight. *Proceedings of the American Institute of Aeronautics and Astronautics Space 2004 conference Exhibi*t. San Diego, CA, Sept.

Bainbridge, W. S. (2010). *Online multiplayer games*. San Rafael, CA: Morgan & Claypool Publishers.

Baldwin, T. T & Ford, J. K. (1988) Transfer of training: A review and directions for future research. *Personnel psychology*, *41*, 63-105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

Barger, L. K., Flynn-Evans, E. E., Kubey, A., Walsh, L., Ronda, J. M., Wang, W., Wright, K. P., Jr, & Czeisler, C. A. (2014). Prevalence of sleep deficiency and use of hypnotic drugs in astronauts before, during, and after spaceflight: an observational study. *The lancet neurology, 13*(9), 904-912. https://doi.org/10.1016/S1474-4422(14)70122-X

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612-637. https://doi.org/10.1037//0033-2909.128.4.612

Barratt, M. (2012, May). *Adaptation to space*. [Oral presentation] the Houston human factors and ergonomic society one-day symposium, Houston TX.

Barshi, I. (2015). From Healy's training principles to training specifications: The case of the Comprehensive LOFT. *American journal of psychology*, *128*(2), 219-227. https://doi.org/10.5406/amerjpsyc.128.2.0219

Barshi, I., & Dempsey, D. L. (2018). *ISS training best practices and lessons learned* [Poster Presentation]. NASA's human research program investigator's workshop. Galveston, TX. https://core.ac.uk/download/pdf/141519949.pdf

Barshi, I., Healy, A., Dempsey, D. L., McGuire, K. M, & Landon, L. B. (2017, January) *The effects of long-duration spaceflight on training retention and transfer*. [Poster Presentation]. NASA's human research program investigator's workshop. Galveston, TX.

Barshi, I., Dempsey, D. L., Landon, L. McGuire, K., Healy, A., & Kole. J. A. (2019) . *Effects of long-duration space flight on training retention and transfer*. Internal NASA Human research program (HRP) human factors and behavioral health (HFBP) element report: unpublished.

Barshi, I., & Loukopoulos, L. D. (2012). Training for real world job performance. In A. F. Healy & L. E. Bourne, Jr. (Eds.). *Training cognition: Optimizing efficiency, durability, and generalizability,* pp. 287-306. New York: Psychology Press.

Basner, M., Dinges, D. F., Mollicone, D., Ecker, A., Jones, C. W., Hyder, E. C., Di Antonio, A. Savelev, I., Kan, K., Goel, N., Morukov, B., & Sutton, J. P. (2013). Mars 520-d mission simulation reveals protracted crew hypokinesis and alterations of sleep duration and timing. *Proceedings of the national academy of sciences*, *110*(7), 2635-2640. www.pnas.org/cgi/doi/10.1073/pnas.1212646110

Basner, M., Savitt, A., Moore, T. M., Port, A. M., McGuire, S., Ecker, A. J., Nasrini, J., Mollicone, D. J., Mott, C. M., McCann, T., Dinges, D. F., & Gur, R. C. (2015). Development and validation of the cognition test battery for spaceflight. *Aerospace medicine and human performance*, *86*(11), 942-952. https://doi.org/10.3357/AMHP.4343.2015

Beard, B.L. (2020). *Characterization of how C0$_2$ level may impact crew performance related to HSIA risk.* NASA TM 20205011433, https://ntrs.nasa.gov/citations/20205011433.

Beard, B. L., & Ahumada, A. (2019). *The fine motor skills and cognition test batteries: Normative data and interdependencies*. Presented at the 10th annual international association for the advancement of space safety conference : Space safety is no accident, El Segundo, CA, United States. Retrieved November 13, 2019, from https://ntrs.nasa.gov/search.jsp?R=20190025239

Beaubien, J.M., & Baker, D.P. (2002). A review of selected aviation Human Factors taxonomies, accident/incident reporting systems, and data reporting tools. *International journal of applied aviation studies 2*(2):11-36.

Beer, J., Fisk, A., & Rogers, W. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, *3*(2), 74-99. https://doi.org/10.5898/JHRI.3.2.Beer

Begault, D.R. (2011, October). *Effect of whole-body vibration on speech. Part II: Effect on intelligibility.* Audio engineering society 131st convention: Audio engineering society, New York, NY.

Begault, D. R., Godfroy, M., Holden, K., & Sándor, A. (2008). *Caution and warning alarm design and evaluation for NASA CEV auditory displays*. Paper presented at the audio engineering society 125th convention, San Francisco, https://www.aes.org/events/125/.

Begault, D. R., Godfroy, M., Sándor, A., & Holden, K. (2007). Auditory alarm design for NASA CEV applications. *Proceedings of the 13th international conference on auditory sisplay*, pp. 131-137.

Berger, M., Mescheriakov, S., Molokanova, E., Lechner-Steinleitner, S., Seguer, M., & Kozlovskaya, I. (1997). Pointing arm movements in short- and long-term spaceflights. *Aviation, space, and environmental medicine*, *68*, 781-787.

Beyer, H., & Holtzblatt, K. (1997). *Contextual design: Defining customer-centered systems* (1st ed.). Morgan Kaufmann.

Billings, C. E. (1997). *Aviation automation: The search for a human centered approach*. Mahwah, NJ: Erlbaum.

Billings, D. R., Schaefer, K. E., Chen, J. Y. C., & Hancock, P. A. (2012). Human-robot interaction: Developing trust in robots. *Proceedings of the seventh annual ACM/IEEE international conference on human-robot interaction*. https://doi.org/10.1145/2157689.2157709

Billman, D. & Catrambone, R. (2019). *Training for generalizable skills & knowledge: Integrating principles and procedures* [Unpublished report]. NASA human research program (HRP) human factors and behavioral health (HFBP) element.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *The American psychologist*, *61*(1), 27-41. https://doi.org/10.1037/0003-066X.61.1.27

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic bulletin & review*, *4*(2), 184-197. https://doi.org/10.3758/BF03209393

Blume, B. D., Ford, K. A., Baldwin, T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of management*, *36*, 1065-1105. https://doi.org/10.1177%2F0149206309352880

Bock, O., Fowler, B., & Comfort, D. (2001). Human sensorimotor coordination during spaceflight: An analysis of pointing and tracking responses during the "Neurolab" space shuttle mission. *Aviation, space, and environmental medicine*, 72, 877-883.

Bracken, B., Leather, C., Cross, E. V., Stephenson, J., Greene, M., Lancaster, J., Munson, B., & Holden, K. (2019). Validation of a physiological approach to measure cognitive workload: CAPT PICARD. In L. Longo & M. C. Leva (Eds), *Proceedings of the third international symposium, H-WORKLOAD human mental workload: Models and applications*. https://doi.org/10.1007/978-3-030-32423-0_5

Brainard, G. C., & Hanifin, J. M. (2005). Photons, clocks, and consciousness. *Journal of biological rhythms, 20*(4),314-325. https://doi.org/10.1177/0748730405278951

Brainard, G. C., Hanifin, J. P., Greeson, J. M., Byrne, B., Glickman, G., Gerner, E., & Rollag, M. D. (2001). Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *The journal of neuroscience: the official journal of the society for neuroscience, 21*(16), 6405-6412. https://doi.org/10.1523/JNEUROSCI.21-16-06405.2001

Brainard, G.C., Sliney, D., Hanifin, J. P., Glickman, G., Byrne, B., Greeson, J. M., Jasser, S., Gerner, E., & Rollag, M. D (2008). Sensitivity of the human circadian system to short-wavelength (420-nm) light. *Journal of biological rhythms, 23*(5):379-386. https://doi.org/10.1177/0748730408323089

Bray, T. M., Shehab, R. L., Schlegel, R. E., Civan, A., & Walker, D. E. (2003). Effectiveness of 2-D views for 6-D robotics simulation maneuvers. *Proceedings of the Human factors and ergonomics society annual meeting*, *47*(8), 975-979. https://doi.org/10.1177/154193120304700805

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. London: Taylor and Francis.

Brown, J. (1975). *ASTP002: Skylab 4 and ASTP crew height.* Retrieved December 28, 2007, from NASA life sciences data archive website http://lsda.jsc.nasa.gov

Brown, (1976). *Zero-g effects on crewman height*. JSC Internal Note 76-EW-3. NASA, Houston, TX.

Bualat, M., Fong, T., Schreckenghost, D., Kalar, D., Pacis, E., & Beutter, B. (2014, June 17). *Results from testing crew-controlled surface telerobotics on the International Space*

Station. [Oral presentation] The i-SAIRAS international symposium on artificial intelligence, robotics and automation in space, Montreal, Canada.

Butler, K. A., Zhang, J., Esposito, C., Bahrami, A., Hebron, R., & Kieras, D. (2007). Work-centered design: A case study of a mixed-initiative scheduler. *Proceedings of the SIGCHI conference on human factors in computing systems*, 747-756. ACM.

Cabibihan, J., Wing-Chee So, & Pramanik, S. (2012). Human-recognizable robotic gestures. *IEEE transactions on autonomous mental development*, *4*(4), 305-314. https://doi.org/10.1109/TAMD.2012.2208962

Calhoun, G. L., & Draper, M. H. (2015). Display and control concepts for multi-UAV applications. In K. P. Valavanis & G. P. Vachtsevanos (Eds.), *Handbook of unmanned aerial vehicles*, pp. 2443-2473. https://doi.org/10.1007/978-90-481-9707-1_88

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage Publications.

Carney, K. (2015). *50th percentile male human vibration model for coupled loads analysis* [Unpublished report]. NESC for work performed under NESC Assessment TI-14-00936.

Casey, S. M. (1993). *Set phasers on stun: And other true tales of design, technology, and human error*. Aegean: Santa Barbara, CA.

Casler, J. G., & Cook, J. R. (1999). Cognitive performance in space and analogous environments. *International journal of cognitive ergonomics*, *3*(4), 351-372. https://doi.org/10.1207/s15327566ijce0304_5

Casner, S. M., & Gore, B. F. (2010). *Measuring and evaluating workload: A primer.* NASA technical memorandum, NASA/TM 2010- 216395. https://matb-files.larc.nasa.gov/Workload_Primer_TM_Final.pdf

Casper, J., & Murphy, R. R. (2003). Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE transactions on systems, man, and cybernetics—part b cybernetics*, *33*(3), 367-385. https://doi.org/10.1109/TSMCB.2003.811794

Chadwick, R. A., Pazuchanics, S. L., & Gillan, D. J. (2006). What the robot's camera tells the operator's brain. Cooke, N.J., Pringle, H.L., Pedersen, H.K. and Connor, O. (Ed.s) *Advances in human performance and cognitive engineering research*: *vol. 7. human factors of remotely operated vehicles*, *7*, 373-384. http://dx.doi.org/10.1016/S1479-3601(05)07026-8

Chandler, F. (2007). *Learning from NASA mishaps: What separates success from failure*. Report presented to the project management Challenge, February 7, 2007. Available from: http://spacese.spacegrant.org/Failure/Reports/Chandler_Faith_LearningFromFailures.pdf

Chang, M. L., & Marquez, J. J. (2018). *Human-automation allocations for current robotic space operations: space station remote manipulator system.* NASA technical memorandum, NASA/TM—2018–220042. https://ntrs.nasa.gov/citations/20200003149

Chen, J. Y. C., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on muman-machine systems*, *44*(1), 13-29. https://doi.org/10.1109/THMS.2013.2293535

Chen, J. Y. C., & Joyner, C. (2009). Concurrent performance of gunner's and robotics operator's tasks in a multi-tasking environment. *Military psychology*, *21*, 98-113. https://doi.org/10.1080/08995600802565785

Chen, M, Arellano, J., Meyers, J., Hsiang, S., Lim, C., Ramirez, C., Alaimo, R., & Stearns, J. (2020). *The spacecraft optimization layout and volume (SOLV) extension task final report.* [Unpublished report]. NASA human research program (HRP) human factors and behavioral health (HFBP) element.

Chiles, J. R. (2002). *Inviting disaster: Lessons from the edge of technology: An inside look at catastrophes and why they happen*. New York: Harper Paperbacks.

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. *Advances in human performance and cognitive engineering research*, *2*, 1-12. https://doi.org/10.1016/S1479-3601(02)02003-9

Churchill, E., Laubach, L.L., McConville, J.T., & Tebbetts, I. (1978). *Anthropometric source book, vol 1: Anthropometry for designers.* NASA, Washington, D.C.

Clément, G. (2007). Using your head: Cognition and sensorimotor functions in microgravity. *Journal of the American society for gravitational and space research, 20*(2).

Cohen, M., Parasuraman, R., & Freeman, J. (1998). Trust in decision aids: A model and its training implications. *Proceedings 1998 command and control research and technology symposium CCRP,* Washington, DC. http://www.cog-tech.com/papers/Trust/C2_Trust_paper_revised.pdf

Coleshill, E., Oshinowo, L., Rembala, R., Bina, B., Rey, D., & Sindelar, S. (2009). Dextre: Improving maintenance operations on the International Space Station. *Acta astronautica*, *64*, 869-874. https://doi.org/10.1016/j.actaastro.2008.11.011

Conners, M.M., Harrison, A.A., & Akins, F.R. (1986). Psychology and the resurgent space program. *American psychologist, 41*(8), 906-913. https://doi.org/10.1037/0003-066X.41.8.906

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. *Human factors: The journal of the Human Factors and Ergonomics Society*, *42*(1), 151-173. https://doi.org/10.1518/001872000779656561

Cooper, G. E., & Harper Jr., R. P. (1969). The use of pilot rating in the evaluation of aircraft handling qualities. NASA technical note, NASA-TN-D-5153. NASA. https://ntrs.nasa.gov/search?q=use%20of%20pilot%20rating

Crandall, J. W., & Goodrich, M. A. (2001). Experiments in adjustable autonomy. *Systems, man, and cybernetics, 2001 IEEE international conference on systems, man and cybernetics*, *3*, 1624-1629. https://doi.org/10.1109/ICSMC.2001.973517

Cristoforetti, S. [@AstroSamantha]. (2014, June 28) Can't train that on Earth, of course, but pushing off the "ceiling" with feet is one way to do #CPR in space. [Tweet] https://twitter.com/astrosamantha/status/482849360871309314

Cullen, W. (2001). *The Ladbroke Grove rail inquiry - part 2 report.* London: HSE Books.

Cummings, M. L. (2004). Automation bias in intelligent time critical decision support systems. *Proceedings of the AIAA 3rd intelligent systems conference*. https://doi.org/10.2514/6.2004-6313

Cummings, M. L. (2014). Man versus machine or man + machine? *IEEE intelligent systems*, *29*(5), 62-69. https://doi.org/10.1109/MIS.2014.87

Cummings, M. L. (2017). Informing autonomous system design through the lens of skill-, rule-, and knowledge-based behaviors. *Journal of cognitive engineering and decision making*, 12(1), 58-61. https://doi.org/10.1177/1555343417736461

Cummings, M. L. (2019a). Adaptation of licensing examinations to the certification of autonomous systems. In X. Li, R. Murray, C. J. Tomlin & H. Yu (Eds.), *Safe, autonomous and intelligent vehicles* (1st ed.). Springer.

Cummings, M. L. (2019b). Global order and the (mis)perception of powerful AI. *Proceedings of the 2019 global order colloquium.* http://hal.pratt.duke.edu/sites/hal.pratt.duke.edu/files/u35/misperceptionAI%2026AUG.pdf

Currie, N. J., & Peacock, B. (2002). International Space Station robotic systems operations—A human factors perspective. *Proceedings of the human factors and ergonomics society annual meeting*, *46*(1), 26-30. https://doi.org/10.1177/154193120204600106

Currie, N., & Rochlis, J. (2004). Command and telemetry latency effects on operator performance during International Space Station robotics operations. *Proceedings of the human factors and ergonomics society annual meetin*g, 48(1), 66-70. https://doi.org/10.1177/154193120404800115

Da-Poian, V., & Koryanov, V. V. (2018). *C*RAFT: Collaborative rover and astronauts future technology. *Proceedings of the deep space Gateway science workshop.* https://www.hou.usra.edu/ meetings/deepspace2018/pdf/3030.pdf

Darken, R. P., & Cervik, H. (1999). Map usage in virtual environments: Orientation issues. *Proceedings of the 1999 IEEE conference on virtual reality*, pp. 133-140. doi: 10.1109/VR.1999.756944.

Dautenhahn, K. (2007). Methodology and themes of human-robot interaction: A growing research field. *International Journal of advanced robotic systems*, *4*(1), 103-108. https://doi.org/10.5772/5702

De Barros, P. G., & Lindeman, R. W. (2008). A survey of user interfaces for robot teleoperation. Technical report No. WPI-CS-TR-09-12. Worcester Polytechnic Institute. https://core.ac.uk/download/pdf/47187275.pdf

de Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International journal of human-computer studies*, *58*, 719-753. https://doi.org/10.1016/S1071-5819(03)00039-9

de Winter, J., & Dodou, D. (2011). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, technology & work*, *16*, 1-11. https://doi.org/ 10.1007/s10111-011-0188-1

de Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour*, 27, 196-217.

Defazio, P.A. & Larson, R. (2020). *The design, development & certification of the Boeing 737 MAX.* The House Committee on Transportation & Infrastructure. Washington, D.C. The House Committee on Transportation and Infrastructure. https://transportation.house.gov/

imo/media/doc/2020.09.15%20FINAL%20737%20MAX%20Report%20for%20Public%20Release.pdf

Defense Science Board. (2016). *Defense science board summer study of autonomy*. Washington, DC: Under Secretary of Defense.

Degani, A. (2001). *Korean Air Lines flight 007: Lessons from the past and insights for the future.* Ames Research Center, NASA. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.3565&rep=rep1&type=pdf

Degani, A. (2004). *Taming HAL: Designing interfaces beyond 2001.* New York: Palgrave MacMillan.

Degani, A., & Heymann, M. (2002). Formal verification of human-automation interaction. *Human factors, 44*(1), 28-43.

DeJong, B. P., Colgate, J. E., & Peshkin, M. A. (2004). Improving teleoperation: Reducing mental rotations and translations. *Proceedings of the IEEE International conference on robotics and automation, 14*, 3708-3714. doi: 10.1109/ROBOT.2004.1308838

Dempsey, D. L., & Barshi, I. (2020). Applying research-based training principles: Towards crew-centered, mission-oriented space flight training. In L.B. Landon, K.J. Slack & E. Salas (Eds.), *Psychology and human performance in space programs*, pp. 63-80. Abingdon, UK: Taylor & Francis.

Dempsey, D. L., Barshi, I., Landon, L. B., Jensen, M, Fard, B., Nyberg, K., & Hansen, J. (2018). *Training for NASA's Gateway and deep space transport (DST) missions: FOD stakeholder concerns and top priorities workshop* [Unpublished report]. NASA human research program (HRP) human factors and behavioral health (HFBP) element.

Dennehy, K. 1997. C*ranfield – Situation awareness scale, user manual*. Applied psychology unit, College of Aeronautics, Cranfield University, COA report No. 9702.

Desai, M., Medvedev, M., Vazquez, M., McSheehy, S., Gadea-Omelchenko, S., Bruggeman, C., Steinfeld, A., & Yanco, H. (2012). Effects of changing reliability on trust of robot systems. *Proceedings of the 7th ACM/IEEE international conference on human-robot interaction* (HRI), Boston, MA, 2012, pp. 73-80. doi: 10.1145/2157689.2157702.

DeSantis, L., Archer, R., Perez, E., Blackledge, C., Wong, D., & Rajulu, S. (2011). J*oint EVA NBL Orion mock-up (JENOM) evaluation human factors data analysis.* Johnson Space Center.

Desaulniers, D. R., Gillan, D. J., & Rudisill, M. (1988). The effects of format in computer-based procedure displays. *Proceedings of the human factors and ergonomics society 32nd annual meeting*, pp. 291-295

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological science*, *23*(12), 1549-1556. https://doi.org/10.1177/0956797612448793.

de Visser, E., Kidwell, B., Payne, J., Lu, L., Parker, J., Brooks, N., Chabuk, T., Spriggs, S., Freedy, A., Scerri, P., & Parasuraman, R. (2013). Best of both worlds: Design and evaluation of an adaptive delegation interface. *Proceedings of the human factors and ergonomics society annual meeting*, *57*(1), 255–259. https://doi.org/10.1177/1541931213571056

Di Nocera, F., Lorenz, B., & Parasuraman, R. (2005). Consequences of shifting from one level of automation to another: Main effects and their stability. In D. de Waard, K. A. Brookhuis, R. van Egmond & T. Boersema (Eds.), *Human Factors in design, safety, and management*, pp. 363-376. Maastricht, the Netherlands: Shaker Publishing.

DiLaura, D., Houser, K., Mistrick, R., & Steffy, G. (2011). *The lighting handbook*, 10th ed. Illuminating engineering society, New York, New York.

Diftler, M., Badger, C., Potter, E., & Pike, L. (2015). Robonaut 2—Building a robot on the International Space Station. *Proceedings of 2015 ISS research and development conference*. Presented at the Boston, MA. Boston, MA.

Dirlich, T. (2010). Experiments in simulated microgravity under water: preliminary findings on individual neutral postures. In Duffy, V. (Ed.), *Advances in applied digital human modeling.* Taylor & Francis group. https://doi.org/10.1201/EBK1439835111

Dismukes, R. K., Berman, B., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents*. Burlington, VT: Ashgate.

Dix, A., Finlay, J., Abowd, G., & R. Beale. (2004). *Human computer interaction, 3rd Edition*. Pearson Education Limited: Essex, England.

Dixon, K. R., & Khosla, P. K. (2004). Learning by observation with mobile robots: A computational approach. *Proceedings of the IEEE international conference on robotics and automation, 1,* 102-107. https://doi.org/10.1109/ROBOT.2004.1307136

Drury, J., Keyes, B., & Yanco, H. (2007). LASSOing HRI: analyzing situation awareness in mapcentric and video-centric interfaces. *Proceedings of the second ACM SIGCHI/SIGART conference on human-robot interaction*, pp. 279-286. DOI: 10.1145/1228716.1228754

Duda, K. R., Prasov, Z., York, S. P., West, J. J., Robinson, S. K., & Handley, P. M. (2015). Development of an integrated simulation platform for real-time task performance assessment. *Proceedings of the IEEE aerospace conference*, pp. 1-9. https://doi.org/10.1109/AERO.2015.7118974

Durso, F.T. , & Dattel, A.R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application*, pp. 137-154. Aldershot, UK: Ashgate.

Dutch Safety Board. (2010). *Crashed during approach, Boeing 737-800, near Amsterdam Schipol Airport, 25 February 2009*. https://www.onderzoeksraad.nl/en/page/ 1182/turkish-airlines-crashed-during-approach-boeing-737-800-amsterdam

EC/Chief Crew Systems Division. (1968). *Memorandum report on X axis manned vibration tests from Flight 502 and possible amplification factors of vibratio*n. In: PA/Manager Apollo spacecraft program (ed). Houston, TX. https://www.ibiblio.org/apollo/Documents/lvfea-AS502-Apollo6.pdf

Ellis, S.R. (2000). Collision in Space. *Ergonomics in design: The quarterly of human factors applications*, *8*(1), 4-9. https://doi.org/10.1177/106480460000800102

Ellis, S.R., & Liston, D. B. (2011). Static and motion-based visual features used by airport tower controllers: Some implications for the design of remote or virtual towers. NASA Technical report, NASA/TM-2011-216427. https://ntrs.nasa.gov/citations/20140000575

Ellis, Stephen R., & Adelstein, B. D. (2014). Interaction of movement difficulty due to the misalignment effect and response latency: Connections to Fitts' law and implications for system latency requirements. *Proceedings of the human factors and ergonomics society annual meeting*, *58*(1), 1756-1760. https://doi.org/10.1177/1541931214581366

Embrey, D., Blackett, C., Marsden, P, & Peachey, J. (2006). *Development of a human cognition workload assessment tool.* Human reliability associates. http://citeseerx.ist.psu.edu/ viewdoc/download?doi=10.1.1.138.4767&rep=rep1&type=pdf

Endsley, M. R. (1988a). Design and evaluation for situation awareness enhancement. *Proceedings of the human factors society 32nd annual meeting*, *32*(2), 97-101. https://doi.org/ 10.1177%2F154193128803200221

Endsley, M.R. (1988b). Situation awareness global assessment technique (SAGAT). *Proceedings of the IEEE 1988b national aerospace and electronics conference*, pp. 789-795 vol.3, doi: 10.1109/NAECON.1988.195097.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human factors, 37*(1), 65-84. https://doi.org/10.1518%2F001872095779049499

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human factors*, *37*, 32-64. https://doi.org/10.1518%2F001872095779049543

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: a critical review. In M. R. Endsley & D. J. Galrnad (Eds), *Situation awareness analysis and measurement*, pp. 3-32, CFC Press.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human-automation research. *Human factors*, *59*(1), 5-27. https://doi.org/10.1177/0018720816681350.

Endsley, M., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492. https://doi.org/10.1080/001401399185595

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, *37*, 381-394. https://doi.org/10.1518%2F001872095779064555

Energia (2004). NASA space science coordinated archive, NSSDCA/COSPAR ID:2004-040A. https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=2004-040A

England, S., Benson, E., & Rajulu, S. (2008). Functional mobility testing: A novel method to establish human-system interface design requirements. *Proceedings of the ASME 2008 Summer bioengineering conference. ASME 2008 Summer bioengineering conference, parts A and B*, pp. 907-908. https://doi.org/10.1115/SBC2008-192925

Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human factors, 41*(2), 312–325. https:// https://doi.org/10.1518/001872099779591196

Eppler, D., Evans, C., Tewskbury, B., Helper, M., Bleacher, J., Fossum, M., Ross, D., & Feustel, A. (2016). Geologic training for America's astronauts. *GSA Toda*y, *26*(8). DOI: 10.1130/GSATG295GW.1.

Evans, B (2011). *At home in space: the late seventies into the eighties.* Springer.

Ezer, N. (2011). Human interaction within the “glass cockpit”: Human engineering of Orion display formats. *Proceedings of the 18th IAA humans in space symposium* (#2324).

Ezer, N., Sandor, A., & Vos, G. (2013). *Human-robot interaction (HRI): Metrics and methods* [Unpublished Report]. NASA Johnson Space Center.

Fan, T., Hyams, D., & Kuchar, J. K. (1998). *Preliminary study of in-flight replanning performed on the flight deck*. National center of excellence in aviation operations research (NEXTOR).

Feary, M., McCrobie, D., Alkin, M., Sherry, L., Polson, P., Palmer, E., & McQuinn, N. (1998). *Aiding vertical guidance understanding*. NASA technical report, NASA/TM-1998-112217. https://ntrs.nasa.gov/citations/19980018859

Federal Aviation Administration. (1996). *The interface between flight crews and modern flight deck systems*. Federal Aviation Administration. http://www.faa.gov/aircraft/air_cert/design_approvals/csta/publications/media/fltcrews_fltdeck.pdf

Feigh, K. M., & Pritchett, A. R. (2013). Requirements for effective function allocation: A critical review. *Journal of cognitive engineering and decision making, 8*(1), 23-32. https://doi.org/10.1177/1555343413490945

Ferguson, S., Luders, B., Grande, R. C., & How, J. P. (2015). Real-time predictive modeling and robust avoidance of pedestrians with uncertain, changing intentions. In H. L.Akin, N. M. Amato, V. Isler & A. F. van der Stappe (Eds.), *Algorithmic foundations of robotics XI.*, *107*, 161-177. Springer International Publishing.

Ferketic, J., Goldblatt, L., Hodgson, E., Murray, S., Wichowski, R., Bradley, A., Fong, T., Evans, J., Chun, W., Stiles, R., Goodrich, M. A., Steinfeld, A., King, D., & Erkorkmaz, C. (2006). Toward human-robot interface standards II: An examination of common elements in human-robot interaction across the space enterprise. *Proceedings of space 2006*. AIAA 2006-7388. https://doi.org/10.2514/6.2006-7388

Fiore, S. M., Wiltshire, T. J., Lobato, E. J. C., Jentsch, F. G., Huang, W. H., & Axelrod, B. (2013). Toward understanding social cues and signals in human-robot interaction: Effects of robot gaze and proxemic behavior. *Frontiers in psychology*, *4,* 859. https://doi.org/10.3389/fpsyg.2013.00859

Fitts, P. M. (Ed.). (1951). *Human engineering for an effective air-navigation and traffic-control system*. Air navigation development board. https://apps.dtic.mil/dtic/tr/fulltext/u2/b815893.pdf

Fitts, P.M. & Jones, R.E. (1947). *Analysis of factors contributing to 460 "pilot-error" Experiences in operating aircraft controls.* Memorandum report TSEAA-694-12. Aero medical laboratory. Air material command. http://collections.uakron.edu/digital/ collection/p15960coll1/id/25524/

Flemisch, F., Adams, C., Conway, S., Goodrich, K., Palmer, M., & Schutte, P. (2003). T*he h-metaphor as a guideline for vehicle automation and interaction*. NASA technical . memorandum, NASA/TM-2003-212672. NASA. http://ntrs.nasa.gov/archive/nasa /casi.ntrs.nasa.gov/20040031835.pdf

Flentge, F., Weber, S. G., Behring, A., & Ziegert, T. (2008, April). *Designing context- aware HCI for collaborative emergency management*. International workshop on HCI for emergencies in conjunction with CHI (Vol. 8). Florence, Italy.

Fong, T. (2018, August). *Autonomy and the Gateway: Human-robotic systems and operations*. [Oral presentation] Autonomous Crew Operations Technical interchange meeting, NASA Ames Research Center. https://ntrs.nasa.gov/api/citations/ 20180006631/ downloads/20180006631.pdf

Fong, T., & Nourbakhsh, I. (2005). Interaction challenges in human-robot space exploration. *Interactions*, *12*(2), 42-45. https://doi.org/10.1145/1052438.1052462

Fong, T., Micire, M., Morse, T., Park, E., Provencher, C., To, V., Wheeler D. W., Mittman, D., Torres, R. J., & Smith, E. (2013). Smart SPHERES: a telerobotic free-flyer for intravehicular activities in space. *Proceedings of the AIA*A *SPACE 2013 conferences & exposition*, (AIAA 2013-5338). http://dx.doi.org/ 10.2514/6.2013-5338

Fong, T., Nourbakhsh, I., Kunz, C., Fluckiger, L., Schreiner, J., Ambrose, R., Burridge, R., Simmons, R., Hiatt, L., Schultz, A., Trafton, J. G., Bugajska,M., Scholtz, J. (2005). The peer-to-peer human-robot interaction project*. Proceedings of space 2005*, AIAA2005-6750. https://doi.org/10.2514/6.2005-6750

Fong, T., Rochlis Zumbado, J., Currie, N., Mishkin, A., & Akin, D. L. (2013). Space telerobotics: Unique challenges to human-robot collaboration in Space. *Reviews of human factors and ergonomics*, *9*(1), 6-56. https://doi.org/10.1177/ 1557234X13510679.

Ford, J. K., Baldwin, T. T., & Prasad, J. (2018). Transfer of training: The known and unknown. *Annual review of organizational psychology and organizational behavior, 5*, 201-225. https://doi.org/10.1146/annurev-orgpsych-032117-104443

Ford, J. K., & Weissbein, D.A. (1997) Transfer of training: An updated review and analysis. *Performance improvement quarterly, 10*(2):22-41. https://doi.org/10.1111/j.1937-8327.1997.tb00047.x

Forman, R. E. (2011). *Objective performance metrics for improved space telerobotics training*. Thesis (S.M.), Massachusetts Institute of Technology, department of aeronautics and astronautics. http://hdl.handle.net/1721.1/68408

Fothergill, D., Hedges, D., & Morrison, J. (1991). Effects of CO_2 and N_2 partial pressures on cognitive and psychomotor performance. *Undersea biomedical research, 18*(1), 1-19.

Gacy, A. M., Wickens C. D., Sebok, A., Gore, B. F., & Hooey, B L. (2011). Modeling operator performance and cognition in robotic missions. *Proceedings of the 55th annual meeting of the human factors and ergonomics society*, 55(1). https://doi.org/10.1177/1071181311551179

Gallagher, W. (1993). *The power of place*. Poseidon Press, Simon & Schuster Inc., New York.

Gawande, A. (2009). *The checklist manifesto: How to get things right.* NewYork: Henry, Holt and Company.

Gawron, V. J. (2000). *Human performance measures handbook*. CRC Press.

Geiger, L., Popp, M., Färber, B., Artigas, J., & Kremer, P. (2010). The influence of telemanipulation-systems on fine motor performance. *Proceedings of the 2010 third international conference on advances in computer-human interactions*, 44-49. https://doi.org/10.1109/ACHI.2010.46

Gervits, F., Fong, T. W., & Scheutz, M. (2018). Shared mental models to support distributed human-robot teaming in space. Proceedings of the *AIAA SPACE and astronautics forum and exposition*. AIAA 2018-5340. https://doi.org/10.2514/6.2018-5340

Glover, S. M., Prawitt, D. F., & Spilker, B. C. (1997). The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational behavior and human decision processes*, *72*, 232-255. https://doi.org/10.1006/obhd.1997.2735

Godinez, A., Liston, D. B., Ayzenberg, R., Toscano, W. B., Cowings, P. A., & Stone, L.S. (2014). G-loading and vibration effects on heart and respiration rates. *Aviation, space, and environmental medicine, 85*(9), 949-953. https://doi.org/10.3357/ASEM.4015.2014

Gonzalez, L. J., Maida, J., Miles, E. H., Rajulu, S., & Pandya, A. K. (2002). Work and fatigue characteristics of unsuited and suited humans during isolated isokinetic joint motions. *Ergonomics, 45*(7), 484-500. https://doi.org/10.1080/00140130210139582

Goodman, J. R. (2003). International Space Station acoustics. *Proceedings of the NOISE-CON 2003*. https://www.nasa.gov/centers/johnson/pdf/486038main _InternationalSpaceStationAcoustics.pdf

Gooley, J. J., Rajaratnam, S. M., Brainard, G. C., Kronauer, R. E., Czeisler, C. A., & Lockley, S. W. (2010). Spectral responses of the human circadian system depend on the irradiance and duration of exposure to light. *Science translational medicine, 2*(31):31-33. doi: 10.1126/scitranslmed.3000741

Gore, B. F. (2010). Man-machine integration design and analysis system (MIDAS) v5: Augmentations, motivations, and directions for aeronautics applications. In P. C. Cacciabu, M. Hjalmdahl, A. Luedtke & C. Riccioli (Eds.), *Human modeling in assisted transportation*, pp. 43-54. Heidelberg: Springer, NY.

Gore, B.F., Casner, S.M, & Wickens, C.D. (2018). Workload and human performance. In B. Kanki, J-F Clervoy, & G. Sandal (Eds.), *Space safety and human performance*, pp.54-70. Oxford, MA: Elsevier Science.

Green, M., Thaxton, S., & Adolf, J. (2018). *Habitability assessment of International Space Station (ISS Habitability)-final report* [Unpublished report]. NASA's Human Research Program.

Gresty, M. A., Golding, J. F., Le, H., & Nightingale, K. (2008). Cognitive impairment by spatial disorientation. *Aviation, space, and environmental medicine*, *79*(2), 105–111.

Griffin, M. J. (1990). *Handbook of human vibration*. Academic Press, London.

Grosveld, F. W., Goodman, J. R., & Pilkinton, G. D. (2003). International Space Station acoustic noise control-case studies. *Proceedings of the 2003 national conference on noise control engineering.* http://www.grosveld.com/publications/NoiseCon2003case.pdf

Grote, G. (2009). *Management of uncertainty: Theory and application in the design of systems and organizations*. London; Springer.

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human-robot teams. *Interaction studies*, *8*(3), 483-500. https://doi.org/10.1075/is.8.3.10gro

Halloran, J., Fitzpatrick, G., Rogers, Y., & Marshall, P. (2004, April). Does it matter if you don't know who's talking?: multiplayer gaming with voiceover IP. *In extended abstracts on Human factors in computing systems,* pp. 1215-1218. https://doi.org/10.1145/ 985921.986027

Hamilton, D. (2008). Cardiovascular disorders. In M. R. Barratt & S. L. Pool (Eds.), *Principles of clinical medicine for space flight,* pp. 317-359. Springer.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, *53*(5), 517-527. https://doi.org/10.1177/0018720811417254

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati, N. (Eds.), *Advances in psychology*: *Volume 52*, 139-183. http://www.sciencedirect.com/ science/article/pii/S0166411508623869

Head, J. W., Pieters, C. M., & Scott, D. R. (2018, Feb). *Deep space gateway facilitates exploration of planetary crusts: A human/robotic exploration design reference campaign to the lunar Orientale basin.* [Oral presentation] Deep space Gateway science workshop, Denver, CO. https://www.hou.usra.edu/meetings/ deepspace2018/ pdf/3157.pdf

Healy, A. F. (2007). Transfer: Specificity and generality. In H. L. Roediger, III, Y. Dudai & S. M. Fitzpatrick (Eds.), *Science of memory: Concepts*, pp. 271-275. New York: Oxford University Press.

Healy, A. F., & Bourne, L. E., Jr. (1995). Durability and specificity of knowledge and skills. In A. F. Healy & L. E. Bourne, Jr. (Eds). *Learning and memory of knowledge and skills: Durability and specificity*, pp. ix-xiv. Thousand Oaks, CA: Sage Publications.

Healy, A. F., & Bourne, L. E., Jr. (Eds.) (2012). *Training cognition: Optimizing efficiency, durability, and generalizability*. New York: Psychology Press.

Healy, A. F., Wohldmann, E. L., & Bourne, L. E., Jr. (2005). The procedural reinstatement principle: Studies on training, retention, and transfer. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications*, pp. 59-71. Washington, DC: American Psychological Association.

Heikkilä, S. S., Halme, A., & Schiele, A. (2012). Affordance-based indirect task communication for astronaut-robot cooperation. *Journal of field robotics*, *29*(4), 576-600. https://doi.org/10.1002/rob.21413

Helmreich, R. L., & Merrit, A. C. (1998). *Culture at work in aviation and medicine: National, organizational, and professional influences*. Burlington, VT: Ashgate.

Heymann, M., Degani, A., & Barshi, I. (2007). Generating procedures and recovery Sequences: A formal approach. *Proceedings of the 14th international symposium of aviation psychology*. https://corescholar.libraries.wright.edu/isap_2007

Hill, S. G., Iaveccia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human factors, 34*, 429-439. https://doi.org/10.1177/001872089203400405

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The journal of the human factors and ergonomics society*, *57*(3), 407-434. https://doi.org/10.1177/0018720814547570

Hoffmann, E. R. (1992). Fitts' Law with transmission delay. *Ergonomics*, *35*(1), 37-48. https://doi.org/10.1080/00140139208967796

Hoffman, R., Feltovich, P., Fiore, S., Klein, G., Missildine, W., & DiBello, L. (2010). *Accelerated proficiency and facilitated retention: Recommendations based on an integration of research and findings from a working meeting.* Air Force Research Lab. Mesa, AZ.

Hogervorst, M. A., Brouwer, A. M., & Van Erp, J. B. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience, 8*, 322. https://doi.org/10.3389/fnins.2014.00322

Holden, K., Greene, M., Cross, E.V., & Feiveson, A. (2018a). *Effects of long-duration microgravity on fine motor skills*. [Oral presentation] the human research program (HRP) investigator's workshop, Galveston, TX.

Holden, K., Munson, B., Russi-Vigoya, N., Dempsey, D., & Adelstein, B. D. (2019a). *Human capabilities assessment for autonomous missions (HCAAM) phase II: Development and validation of an autonomous operations task list* [Unpublished Report]. NASA JSC.

Holden, K., Munson, B., Russi-Vigoya, M.N., Dempsey, D., & Adelstein, B. (2020). *Human Capabilities Assessment for Autonomous Missions (HCAAM) Phase II: Development and Validation of an Autonomous Operations Task List Final Report*. [Unpublished Report]. NASA human research program (HRP) human factors and behavioral health (HFBP) element.

Holden, K., Russi-Vigoya, N., Adelstein, B. D., & Munson, B. (2019b). *Human capabilities assessment for autonomous missions (HCAAM) phase I: Human performance standards and guidelines final report* [Unpublished Report]. NASA JSC.

Holden, K., Schreckenghost, D, Greene, M, Cross, E.V., Lancaster, J. Stephenson, J., & Morin, L. (2018b January). *Electronic procedure design for crewed missions beyond low Earth orbit (LEO).* [Oral presentation] NASA human research program (HRP) investigator's workshop, Galveston, TX.

Holden, K., Schreckenghost, D., Greene, M., Hamblin, C., Cross II, E. V., Stephenson, J., Lancaster, J., & Morin, L. (2018c). *Electronic procedures for crewed missions beyond low Earth orbit* [Unpublished report] NASA human research program (HRP) human factors and behavioral health (HFBP) element.

Hollnagel, E., & Bye, A. (2000). Principles for modelling function allocation. *International journal of human-computer studies*, *52*(2), 253-265. https://doi.org/10.1006/ijhc.1999.0288

Honn, K. A., Hinson, J. M., Whitney, P., & Dongen, H. P. A. V. (2018). Cognitive flexibility: A distinct element of performance impairment due to sleep deprivation. *Accident analysis and prevention*, *126,* 191-197. https://doi.org/10.1016/j.aap.2018.02.013

Hubens, G., Coveliers, H., Balliu, L., Ruppert, M., & Vaneerdeweg, W. (2003). A performance study comparing manual and robotically assisted laparoscopic surgery using the da Vinci system. *Surgical endoscopy and other interventional techniques*, *17*(10), 1595-1599. https://doi.org/10.1007/s00464-002-9248-1

Hudy, C., Byrne, V., Smith, D., & Whitmore, M. (2005). Human Factors Assessment of Respiratory Support Pack (RSP) Cue Card. *Proceedings of the aerospace medical association annual meeting.*

Hudy, C., & Woolford, B. (2008). *Space human factors engineering gap analysis project final report*. NASA. http://ntrs.nasa.gov/ archive/nasa/casi.ntrs.nasa.gov/20070020456_2007018857.pdf.

Huemer, V. A. Hayashi, M. S., Renema, F.M., Elkins, S., McCandless, J. W., & McCann, R. S. (2005, September). Characterizing scan patterns in a spacecraft cockpit simulator: expert vs. novice performance. *Proceedings of the human factors and ergonomics society annual meeting*, 49(1), 83-87. Sage Publications. https://doi.org/10.1177%2F154193120504900119

Hughes, C. (2013). Executive function: Development, individual differences, and clinical insights. In J. L. R. Rubenstein & P. Rakic (Eds.), *Neural circuit development and function in the brain*, pp. 429-445. https://doi.org/10.1016/B978-0-12-397267-5.00062-5.

Hutchinson, K. J., Watenpaugh, D. E., Murthy, G., Convertino, V. A., & Hargens, A. R. (1995) Back pain during 6 degree head-down tilt approximates that during actual microgravity. *Aviation, space, and environmental medicine, 66*(3):256-259.

Hwang, J.-H., Lee, K., & Kwon, D.-S. (2005). The role of mental model and shared grounds in human-robot interaction. *Proceedings of ROMAN 2005. IEEE International workshop on robot and human interactive communication,* 623-628. https://doi.org/10.1109/ROMAN.2005.1513849

Ijtsma, M., Ma, L. M., Pritchett, A. R., & Feigh, K. M. (2017a). Work dynamics of task work and teamwork in function allocation for manned spaceflight operations. 1*9th International Symposium on aviation psychology,* 554-559. https://corescholar.libraries.wright.edu/isap_2017/58 554-559.

Ijtsma, M., Pritchett, A. R., Ma, L. M., & Feigh, K. M. (2017b). Modeling human-robot interaction to inform function allocation in manned spaceflight operations. *Proceeding of the robotics: science and systems conference*. https://doi.org/10.1177%2F1555343419869484

ISO (1999). *Human-centered design processes for interactive systems*. ISO 13407, 13482. https://www.iso.org/standard/21197.html

Jagacinski, R. J., & Flach, J. M. (2003). *Control theory for humans: Quantitative approaches to modeling performance*. Mahwah, NJ: Lawrence Erlbaum Associates. https://doi.org/10.1002/rnc.920

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International journal of cognitive ergonomics*, *4*(1), 53-71. https://doi.org/10.1207/S15327566IJCE0401_04

Jiang, S. D. (2019). A study of initiative decision-making in distributed human-robot teams. *Proceedings of 2019 Third IEEE International conference on robotic computing (IRC)*, 349-356. https://doi.org/10.1109/IRC.2019.00063

Jiang, S. D., & Odom, J. (2018). Toward initiative decision-making for distributed human-robot teams. *Proceedings of the 6th international conference on human-agent interaction*, 286-292. https://doi.org/10.1145/3284432.3284467

John, B. E. & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM transactions on computer-human interaction*, *3*, 320-351. https://doi.org/10.1145/235833.236054

Johnson, M., Bradshaw, J., Feltovich, P., Jonker, C., van Riemsdijk, B., & Sierhuis, M. (2011). The fundamental principles of coactive design: Interdependence must shape autonomy. In M. de Vos, N. Fornara, J. Pitt, & G. Vouros (Eds.), *Coordination, organizations, institutions, and norms in agent systems VI*, *6541*, pp. 172-191. The Institute for Human & Machine Cognition. http://www.ihmc.us/users/mjohnson/papers/Johnson_ 2010_COIN_FundamentalPrincipleofCoactiveDesign.pdf

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space, and environmental medicine, 67*(6), 507-512.

Jones, D. G., & Endsley, M. R.(2000). Can real-time probes provide a valid measure of situation awareness. *Proceedings of the human performance, situation awareness and automation: user centred design for the new millennium conference.*

Jones, E. M. (2012). *The first lunar landing–Communications transcript, Apollo 11 lunar surface journal*. Original posted 1995, Revised Dec. 7th, 2012. http://www.hq.nasa.gov/office /pao/History/alsj/a11/a11.html

Jones, P. (1995). Cooperative work in mission operations: Analysis and implications for computer support. *Computer supported cooperative work, 3*, 103-145.

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process safety progress*, *16*, 126-131.

Kaber, D.B., Onal, E. & Endsley, M.R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human factors and ergonomics in manufacturing, 10*, 409-430. https://doi.org/10.1002/1520-6564(200023)10:4<409::AID-HFM4>3.0.CO;2-V

Kadous, M., Sheh, K., & Sammut, C. (2006). Effective user interface design for rescue robotics. *Proceeds of the 1st ACM SIGCHI/SIGART Conference on human-robot interaction*, 250-257. https://doi.org/10.1145/1121241.1121285

Kanas, N., Salnitskiy, V., Gushin, V., Weiss, D., Grund, E., Flynn, C., Kozerenko, O., Sled, A., & Marmar, C. (2001). Asthenia-Does it exist in space? *Psychosomatic medicine, 63,* 874-880.

Kantowitz, B. H., & Sorkin, R. D. (1983). *Human factors: Understanding people-system relationships*. New York: John Wiley & Sons.

Karasinski, J., Holder, S., & Robinson, S. K. (2019). *Enabling technologies for deep-space human spaceflight: Human-automation and robotic systems trade analysis* [Unpublished report]. NASA human research program, human factors and behavioral performance (HFBP) element.

Kay, A., Liston, P. M., & Cromie, S. (2014). Measuring crew resource management: Challenges and recommendations. *Proceedings of the 11th international conferences on engineering psychology and cognitive ergonomics*, *8532*, 480-490. https://doi.org/10.1007/978-3-319-07515-0_48

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

Keyes, B., Micire, M., Drury, J. L., & Yanco, H. A. (2010). Improving human-robot interaction through interface evolution. In D. Chugo (ed.), *Human-robot interaction*, pp. 183-202. https://doi.org/10.5772/8127

Kieras, D. (1997). A guide to GOMS model usability evaluation using NGOMSL. In M. G. Helander, T. K. Landauer & P. V. Prabhu (Eds), *Handbook of human-computer interaction*, pp. 733-766. https://doi.org/10.1016/B978-044481862-1.50097-2

Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. *CHI '02 extended abstracts on human factors in computing systems*, 576-577. Elsevier. https://doi.org/10.1145/506443.506491

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. T*he Internet and higher education, 8*(1), 13-24. https://doi.org/10.1016/j.iheduc.2004.12.001

Kinney, J., Luria, S., McKay, C., & Ryan, A. (1979). Vision of submariners. *Undersea biomedical research, 6*, 163-173.

Kirkpatrick, A. W., Blaivas, M., Sargsyan, A. E., McBeth, P. B., Patel, C., Xiao, Z., Pian, L., Panebianco, N., Hamilton, D. R., Ball, C. G., & Dulchavsky, S. A. (2013) Enabling the mission through trans-atlantic remote mentored musculoskeletal ultrasound: case report of a portable hand-carried tele-ultrasound system for medical relief missions. *Telemedicine journal and e-health: the official journal of the american telemedicine association, 19*(7), 530-534. https://doi.org/10.1089/tmj.2012.0243

Kirkpatrick, D. L. (1959). Techniques for evaluation training programs. *Journal of the american society of training directors, 13*, 21-26.

Kirlik, A. (1993). Modeling strategic behavior in human-automation interaction: Why an" aid" can (and should) go unused. *Human factors: The journal of the human factors and ergonomics society*, *35*(2), 221-242. https://doi.org/10.1177%2F001872089303500203

Kirwan, B., & Ainsworth, L. K. (1992). *A guide to task analysis*. London ; Washington, DC: Taylor & Francis.

Knight, W. (2017a, July). Biased algorithms are everywhere, and no one seems to care. *MIT technology review*. https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/

Knight, W. (2017b, October). Forget killer robots—Bias is the real AI danger. *MIT technology review*. https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science*, 36, 757-798. https://doi.org/10.1111/j.1551-6709.2012.01245.x

Kole, J. A., Healy, A. F., Schneider, V., & Barshi, I. (2020). Training principles for declarative and procedural tasks. In Landon, L.B., Slack, K.J., and Salas, E. (Eds.), *Psychology and human performance in space programs*, pp. 131-150. Abingdon, UK: Taylor & Francis. https://doi.org/10.1016/j.jarmac.2012.02.002

Kolers, P. A., & Roediger III, H. L. (1984). Procedures of mind. *Journal of verbal learning and verbal behavior, 23*, 425-449. https://doi.org/10.1016/S0022-5371(84)90282-2

Kolling, A., Walker, P., Chakraborty, N., Sycara, K., & Lewis, M. (2016). Human interaction with robot swarms: A survey. *IEEE transactions on human-machine systems*, *46*(1), 9-26. https://doi.org/10.1109/THMS.2015.2480801.

Konz, S., & Johnson, S. (eds). (1999). *Work design: Industrial egronomics.* Holcomb Hathaway Pubs.

Kranz, G. (2000). *Failure is not an option.* New York, New York: Simon & Schuster.

Lackner, J. R., & DiZio, P. (2000). Human orientation and movement control in weightless and artificial gravity environments. *Experimental brain research, 130*(1), 2-26. https://doi.org/10.1007/s002210050002

Ladkin, P. (1996). AA965 Cali accident report. University of Bielefeld. Retreived July 22, 2011, from http://www.skybrary.aero/bookshelf/books/1056.pdf

Lalor, E. C., Kelly, S. P., Finucane, C., Burke, R., Smith, R., Reilly, R. B., & Mcdarby, G. (2005). Steady-state VEP-based brain-computer interface control in an immersive 3D gaming environment. *EURASIP journal on applied signal processing*, 706906. https://doi.org/10.1155/ASP.2005.3156

Landon, L. B., Slack, K. J., & Barrett, J. D. (2018). Teamwork and collaboration in long-duration space missions: Going to extremes. *American psychologist, 73*(4), 563-575. https://doi.org/10.1037/amp0000260

Landon, L. B., Vessey, W. B., & Barrett, J. D. (2016). *Evidence report: Risk of performance and behavioral health decrements due to inadequate cooperation, coordination, communication, and psychosocial adaptation within a team*. NASA. https://humanresearchroadmap.nasa.gov/Evidence/reports/Team.pdf.

Larsen, C.E. (2008). *NASA experience with pogo in human spaceflight vehicles NATO RTO Symposium ATV-152 on Limit-Cycle Oscillations and Other Amplitude-Limited, Self-Excited Vibrations*. NATO R&T, Norway. https://ntrs.nasa.gov/citations/20080018689

Latorella, K. A. (1996). Investigating interruptions: An example from the flight deck. *Proceedings of the human factors and ergonomics society 40th annual meeting,* pp. 249–253. Santa Monica, CA: Human Factors and Ergonomics Society.

Law, J., Van Baalen, M., Foy, M., Mason, S. S., Mendez, C., Wear, M. L., Meyers, V. E., & Alexander, D. (2014). Relationship between carbon dioxide levels and reported headaches on the international space station. *Journal of occupational and environmental medicine, 56*(5), 477-483. https://doi.org/10.1097/JOM.0000000000000158

Lawton, R., & Ward, N. J. (2005). A systems analysis of the ladbroke grove rail crash. *Accident analysis & prevention, 37*(2), 235-244. https://doi.org/10.1016/j.aap.2004.08.001

Lecuyer, A., Lotte, F., Reilly, R. B., Leeb, R., Hirose, M., & Slater, M. (2008). Brain-computer interfaces, virtual reality, and videogames. *Computer, 41*(10), 66-72. DOI: 10.1109/MC.2008.410

Lee, J. D., & See, K. A. (2004). Trust in computer technology: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, J. D., & Seppelt, B. D. (2009). Human factors in automation design. In S. Nof (Ed.), *Springer handbook of automation*, pp. 1615-1642 New York, NY: Springer.

Lee, J., & Moray, N. (1994). Trust, self-confidence, and operator's adaptation to automation. *International journal of human-computer studies*, *40*(1), 153-184. https://doi.org/10.1006/ijhc.1994.1007

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243-1270. https://doi.org/ 10.1080/00140139208967392

Lee, P. (2002). Mars on Earth. *Ad astra, 14*(3), 12-17. https://space.nss.org/wp-content/uploads/Ad-Astra-Magazine-v14n3F1.pdf

Lee, S., Lau, I. Y., Kiesler, S., & Chiu, C.-Y. (2005). Human mental models of humanoid robots. *Proceedings of the 2005 IEEE international conference on robotics and automation*, 2767-2772. https://doi.org/10.1109/ROBOT.2005.1570532

Leeb, R., Friedman, D., Müller-Putz, G. R., Scherer, R., Slater, M., & Pfurtscheller, G. (2007). Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Computational intelligence and neuroscience*, 2007, 79642. https://doi.org/10.1155/2007/79642

Legner, K. (2004). Humans in space & space biology. *Seminars of the United Nations programme on space applications. Selected papers from activities held in 2003*, pp.77-135. United Nations, New York.

Leveson, N. (1995). *SafeWare: System safety and computers*. Reading, Mass.: Addison-Wesley.

Li, H., Wickens, C. D., Sarter, N., & Sebok, A. (2014). Stages and levels of automation in support of space teleoperations. *Human factors*, *56*(6), 1050-1061. https://doi.org/10.1177/0018720814522830

Lilley, S. (2016, August). *Asynchronicity: The near loss of the Apollo-Soyuz test project crew*. [Oral presentation] the NASA safety Center, NASA. https://nsc.nasa.gov/ docs/default-source/system-failure-case-studies/sfcs-2016-07-15-asynchronicity-presentation.pdf?sfvrsn=95a7eff8_2

Liu, H., & Wang, L. (2018). Gesture recognition for human-robot collaboration: A review. *International journal of industrial ergonomics*, *68*, 355-367. https://doi.org/ 10.1016/j.ergon.2017.02.004.

Liston, D.B., Adelstein, B.D., & Stone, L.S. (2014). Onset of positional vertigo during exposure to combined G loading and chest-to-spine vibration. *Aviation, space, and environmental medicine 85*(2):183-186. https://doi.org/10.3357/asem.3784.2014

Losey, D. P., McDonald, C. G., Battaglia, E., & O'Malley, M. K. (2018). A review of intent detection, arbitration, and communication aspects of shared control for physical human-robot interaction. *Applied mechanics reviews*, *70*(1). https://doi.org/10.1115/1.4039145

Loukopoulos, L. D., Dismukes, R. K., and Barshi, I. (2009). T*he multitasking myth: Handling complexity in real-world operations.* Surrey, England: Ashgate Publishing Limited.

Ma, L. M., Ijtsma, M., Feigh, K. M., Paladugu, A., & Pritchett, A. R. (2018). Modelling and evaluating failures in human-robot teaming using simulation. *Proceedings of the 2018 IEEE aerospace conference.* DOI: 10.1109/AERO.2018.8396581

Ma, L. M., Fong, T., Micire, M. J., Kim, Y. K., & Feigh, K. (2018). Human-robot teaming: concepts and components for design. In M. Hutter & R. Siegwart (Eds.), *Field and service robotics*, pp. 649-663. Springer International Publishing.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, *37*, 30-40. https://doi.org/10.1145/176789.176792

Maida, J., Cross, E., & Tran, K. (2004). Nighttime computer-generated lighting analysis for nighttime viewing of SRMS tile scanning on flight day 2. NASA, FUR-GRAF04-004, p 1-24.

Maida, J., Bowen, C., & Pace, J. (2007). Improving robotic operator performance using augmented reality. *51st annual meeting human factors ergonomics society*, *51*(27), 1635-1639. https://doi.org/10.1177/154193120705102703

Manzey, D. & Lorenz, B. (1998). Joint NASA-ESA-DARA study. Part three: Effects of chronically elevated CO_2 on mental performance during 26 days of confinement. *Aviation, space, and environmental medicine, 69*, 506-514.

Manzey, D., Lorenz, B., & Poljakov, V. (1998). Mental performance in extreme environments: Results from a performance monitoring study during a 438-day spaceflight. *Ergonomics, 41*(4), 537-559. https://doi.org/10.1080/001401398186991

Manzey, D., Lorenz, B., Heuer, H., & Sangals, J. (2000). Impairments of manual tracking performance during spaceflight: More converging evidence from a 20-day space mission. *Ergonomics, 43*(5), 589-609. https://doi.org/10.1080/001401300184279

Manzey, D., Lorenz, B., Schiewe, A., Finell, G., & Thiele, G. (1993). Behavioral aspects of human adaptation to space: Analyses of cognitive and psychomotor performance during an 8-day mission. *Clinical investigator, 71*(9), 725-731. https://doi.org/10.1007/BF00209727

Manzey, D., Lorenz, B., Schiewe, A., Finell, G., & Thiele, G. (1995). Dual-task performance in space: Results from a single-case study during a short term space mission. *Human factors, 37*(4), 667-681. https://doi.org/10.1518/001872095778995599

Marcano, L., Aakre, F., Sannerud, R., & Komulainen, T. (2019). Review of simulator training practices for industrial operators: How can individual simulator training be enabled? *Safety science, 115*, 414-424. https://doi.org/10.1016/j.ssci.2019.02.019

Marquez, J. J., Adelstein, B. D., Chang, M. L., Ellis, S. R., Hambuchen, K., & Howard, R. (2017). *Future exploration missions' tasks associated with the risk of inadequate design of human and automation/robotic.* NASA technical memorandum, NASA/TM-2017-219516. https://ntrs.nasa.gov/citations/20180004123

Marquez, J. J., Adelstein, B. D., Ellis, S. R., Chang, M. L., & Howard, R. (2016). Evaluation of human and automation/robotics integration needs for future human exploration missions. *Proceedings of the 2016 IEEE aerospace conference,* pp. 1-9. doi: 10.1109/ AERO.2016.7500580

Maurino, D., Reason, J., Johnston, N., & Lee, R. (1995). *Beyond aviation human factors.* London: Routledge. https://doi.org/10.4324/9781315261652

McBridge, M., & Morgan, S. (2010). *Trust calibration for automated decision aids*. Institute for Homeland Security Solutions. Research brief, Institute for Homeland Security Solutions, Durham, NC. https://journals.sagepub.com/ doi/full/10.1177/ 0018720819829951

McCann, R., Adelstein, B., Beutter, B., Kaiser, M., Root, P., Godfroy, M., Renema, F., Anderson, M., & Verborgh T. (2009). Effects of thrust oscillation vibration on maintaining situation awareness and assuming manual flight control. [Unpublished Report]. SHFE Information Presentation DRP.

McCann, R.S. & McCandless, J. (2003). Human-machine teaming for dynamic fault management in next-generation launch vehicles. *Proceedings of the joint army-navy-nasa-air force 3rd modeling and simulation meeting*. https://www.researchgate.net/ profile/Robert_Mccann2/publication/242163298_HUMAN-MACHINE_ TEAMING_FOR_DYNAMIC_FAULT_MANAGEMENT_IN_NEXT-GENERATION_LAUNCH_VEHICLES/links/54ea306e0cf25ba91c8291a6.pdf

McConnell, D. S., Muchisky, M. M., & Bingham, G. P. (1998). The use of time and trajectory forms as visual information about spatial scale events. *Perception & psychophysics, 60*(7), 1175-1187. https://doi.org/10.3758/BF03206167

McCurdy, M. (2009). Planning tools for Mars surface operations: Human-computer interaction lessons learned. *Proceedings of the 2009 IEEE aerospace conference*, pp. 1-12. https://doi: 10.1109/AERO.2009.4839639.

McCurdy, M., Connors, C., Pyrzak, G., Kanefsky, R., & Vera, A. H. (Eds.). (2006) Breaking through the fidelity barriers: An examination of our current characterization of prototypes and an example of a mixed-fidelity success. *CHI'06: Proceedings of the SIGCHI conference on human factors in computing systems*, pp.1233-1242. https://doi.org /10.1145/1124772.1124959

McDaniel, M. A., Friedman, A., & Bourne, L. E. (1978). Remembering the levels of information in words. *Memory & cognition, 6,* 156-164. https://doi.org/10.3758/BF03197441

McFarland, R. A. (1953). *Human factors in air transportation: occupational health and safety.* McGraw-Hill.

McGuinness, B. (2004). *Quantitative analysis of situational awareness (QUASA): Applying signal detection theory to true/false probes and self-ratings*. Bae systems Bristol (United Kingdom) advanced technology centre.

McGuirl, J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human Factors*, *48*, 656-665. https://doi.org/10.1518/001872006779166334

McIntyre J., Zago M., Berthoz A., Lacquaniti F. (2003). The brain as a predictor: on catching flying balls in zero-G. In J. C. Buckey & J. L. Homick (Eds.), *The neurolab spacelab mission: Neuroscience research in space*, pp. 55-61. Houston, TX: National Aeronautics and Space Administration, Lyndon B. Johnson Space Center.

McWhorter, L. B., & Reed (2010). *Space shuttle entry digital autopilot*. NASA technical report, SP-2010-3408.

Mechtcheriakov, S., Berger, M., Molokanova, E., Holzmueller, G., Wirtenberger, W., & Lechner-Steinleitner, S., De Col, C., Kozlovskaya, I., & Gerstenbrand, F. (2002). Slowing of human arm movements during weightlessness: The role of vision. *European journal of applied physiology, 87*, 576-583. https://doi.org/10.1007/s00421-002-0684-3

Mecklinger, A., Friederici, A., & Gussow, T. (1994). Confinement affects the detection of low frequency events: An event-related potential analysis. *Journal of psychophysiology, 8*, 98-113.

Meshkati, N. (1991). Human factors in large-scale technological systems' accidents: Three Mile Island, Bhopal, Chernobyl. *Industrial crisis quarterly*, *5*(2), 133-154. https://doi.org/10.1177/108602669100500203

Meisner, E., Isler, V., & Trinkle, J. (2008). Controller design for human-robot interaction. *Autonomous robots*, *24*(2), 123-134. https://doi.org/10.1007/s10514-007-9054-7

Menchaca-Brandan, M. A., Liu, A. M., Oman, C. M., & Natapoff, A. (2007). Influence of perspective-taking and mental rotation abilities in space teleoperation. *HRI '07: Proceedings of the ACM/IEEE international conference on human-robot interaction*, pp. 271-278. https://doi.org/10.1145/1228716.1228753

Mesloh, M., England, S., Benson, L., Thompson, S., & Rajulu, S. (2010) The effects of extravehicular activity (EVA) glove pressure on hand strength. *Proceedings of the applied human factors and ergonomic conference*, 50(1). https://doi.org/10.1177/1071181311551288

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63*(2), 81-97. https://doi.org/10.1037/h0043158

Mindell, D. (2006). Human and machine in the history of spaceflight. In S. J. Dick & R. D. Launius (Eds.), *Critical Issues in the history of spaceflight*, pp. 141-162. NASA.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of management*, *36*(4), 876-910. https://doi.org/10.1177/0149206309356804.

Moore, R. J., Ducheneaut, & Nickell, E. (2007). Doing virtually nothing: Awareness and accountability in massively multiplayer online worlds. C*omputer supported cooperative work*, 16, 265-305. https://doi.org/10.1007/s10606-006-9021-4

Moore, S. T., Dilda, V., Morris, T. R., Yungher, D. A., MacDougall, H. G., & Wood, S. J. (2019). Long-duration spaceflight adversely affects post-landing operator proficiency. *Scientific reports*, *9*(1). DOI: 10.1038/s41598-019-39058-9

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior, 16*, 519-533. https://doi.org/10.1016/S0022-5371(77)80016-9.

Mosier, K., & Skitka, L. J. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications*, pp. 201-220. Hillsdale, NJ: Erlbaum.

Mosier, K., Skitka, L. J., Heers, S., & Burdick, M. D. (1998). Automation bias: Decision making and performance in high-tech cockpits. *International journal of aviation psychology*, *8*, 47-63. https://doi.org/10.1207/s15327108ijap0801_3

Mount, F. E., Whitmore, M., & Stealey, S. L. (2003). *Evaluation of neutral body posture on Shuttle Mission STS-57 (SPACEHAB-1)*. NASA technical memorandum. TM-2003-104805. https://ntrs.nasa.gov/citations/20040200967

Mulder, M. J., Keuken, M. C., van Maanen, L., Boekel, W., Forstman, B. U., & Wagenmakers, E-.J. (2013). The speed and accuracy of perceptual decisions in a random-tone pitch task. *Attention, perception, & psychophysics. 75*(5*)*, 1048-1058. https://doi.org/10.3758/s13414-013-0447-8

Mullin, C., Jr. (1960). Some psychological aspects of isolated Antarctic living. *American journal of psychiatry, 117*, 323-325. https://doi.org/10.1176/ajp.117.4.323

Mumm, J., & Mutlu, B. (2011). Human-robot proxemics: Physical and psychological distancing in human-robot interaction. *Proceedings of the 2011 6th ACM/IEEE international conference on human-robot interaction (HRI),* pp. 331-338. https://doi.org/10.1145/1957656.1957786

NASA (2008). *Human space flight gallery* [photos on the Internet; cited 2008 Jan 26]. Available from: http://spaceflight.nasa.gov/gallery/index.html.

NASA (2009a). *Crew seat DTO project documentation.* Retrieved July 22, 2011, 2011, from http://ea.jsc.nasa.gov/webapp/gfedoc/searchdoc.asp?project=crew_seat_dto&curator=susan.reno-1&rno=nathan.wells-1.

NASA (2009b) *NASA human systems integration requirements*. Constellation program human-system integration requirements (HSIR) (CxP 70024), Rev. C. Houston, TX. NASA, Johnson space center.

NASA (2010). Spektr of failure. *NASA system failures case studies, 4*(11). NASA. https://sma.nasa.gov/docs/default-source/safety-messages/safetymessage-2010-11-08-mirprogresscollision.pdf?sfvrsn=17ae1ef8_4

NASA (2011). *NASA space flight human-system standard*. (NASA-STD-3001, Volume 2). Washington, D.C.: NASA.

NASA (2012). *ISS Generic groundrules and constraints part 2: Execute planning* (No. SSP 50261-02, Rev. D). Houston, TX: International Space Station program.

NASA (2014). *Human integration design handbook (HIDH)* (No. NASA/SP-2010-3407/REV1). Washington, D.C.: NASA.

NASA (2015). NASA-STD-3001 volume 2 revision A. Space flight human systems standard, volume 2: human factors, habitability & environmental health. NASA: Washington, D.C. https://www.nasa.gov/hhp/standards

NASA (2019a). Gateway program human-system requirements (HSR). NASA report GP 10017. https://ntrs.nasa.gov/citations/20190029153

NASA (2019b) NASA-STD-3001 volume 2 revision B. Space flight human systems standard, volume 2: Human factors, habitability & environmental health. NASA: Washington, D.C. https://www.nasa.gov/hhp/standards

NASA Goddard Space Flight Center. (2015). Robotic refueling mission. Retrieved May 26, 2015, from http://ssco.gsfc.nasa.gov/rrm_tools.html

NASA Safety Center. (2009, April). *NASA safety center special study: Shuttle software anomaly*. NASA. https://sma.nasa.gov/docs/default-source/safety-messages/safetymessage-2009-05-01-sts126.pdf?sfvrsn=35aa1ef8_4

NASA Safety Center. (2012, January). *Critical software: Good design built right.* NASA. https://nsc.nasa.gov/docs/default-source/system-failure-case-studies/good-design-built-right.pdf?sfvrsn=c6fecf8_4

National Research Council (NRC) (1993). *Workload transition: Implications for individual and team performance.* B. M. Huey & C. D. Wickens, (eds). Washington DC: NAP.

National Research Council (NRC). (2005). *Assessment of options for extending the life of the Hubble space telescope: Final report*. The National Academies Press. http://www.nap.edu/ catalog/11169/assessment-of-options-for-extending-the-life-of-the-hubble-space-telescope

National Research Council. (2007). *Human-system integration in the system development process: A new look*. The National Academies Press. http://www.nap.edu/catalog/ 11893/human-system-integration-in-the-system-development-process-a-new

National Research Council (2011). *Preparing for the high frontier: The role and training of NASA Astronauts in the post-Space Shuttle era*. Washington, D.C.: The National Academies Press.

Newell, A. and Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In Anderson, J. R., editor, Cognitive skills and their acquisition, pages 1–55. Lawrence Erlbaum Associates, Hillsdale, NJ.

Newman, O. (1972). *Defensible space: Crime prevention through urban design.* Macmillan, New York.

Newman, D. J., & Lathan, C. E. (1999). Memory processes and motor control in extreme environments. *IEEE transactions on systems, man, cybernetics - part C (applications and reviews), 29*(3), 387 - 394. doi:10.1109/5326.777074

Nielsen, C. W., Goodrich, M. A., & Ricks, R. W. (2007). Ecological interfaces for improving mobile robot teleoperation. *IEEE transactions on robotics*, *23*(5), 927-941. https://doi.org/10.1109/TRO.2007.907479

Nikolaidis, S., & Shah, J. (2013). Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy. *Proceedings of the 8th ACM/IEEE International conference on human-robot interaction*, 33-40. ACM Digital Library. http://dl.acm.org/citation.cfm?id=2447556.2447563

Norris, J., Powell, M., Vona, M., Backes, P., & Wick, J. (2005). Mars exploration rover operations with the science activity planner. *Proceedings of the* 2005 *IEEE conference on robotics and automation*, pp. 4618-4623. doi: 10.1109/ROBOT.2005.1570832.

Nova, N. (2005). A review of how space affords socio-cognitive processes during collaboration. *PsychnNology journal*, *3*(2), 118-148.

NTSB. (1973). Aircraft accident report- Eastern Air Lines, Inc. L-1011, N310EA, Miami, Florida, December 29, 1972 (No. NTSB-AAR-73-14). Washington, D.C. http://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR73-14.pdf

NTSB. (1986). Aircraft accident report, Galaxy Airlines, Inc. Lockheed Electra-L-188C (No. NTSM/AAR-86/01). Washington, D.C.

NTSB. (1987a). Aircraft accident report - Midair collision of Aeronaves De Mexico, S.A., McDonnell Douglas DC-9-32, XA-JED and Piper PA-28-181, N4891F, Cerritos, California August 31,1986 (No. NTSB/AAR-87/07). Washington, D.C. http://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR86-01.pdf

NTSB. (1987b). Aircraft accident report- Northwest Airlines, Inc. McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne Country Airport, Romulus, Michigan, August 16, 1987 (No. NTSB-AAR-88/05), Washington DC. https://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR88-05.pdf

NTSB. (1991). Aircraft accident report - Avianca, The Airline of Colombia Boeing 707-321 B (No. NTSB/AAR-9 l/04). Washington, D.C. http://libraryonline.erau.edu/online-full-text/ntsb/aircraft-accident-reports/AAR91-04.pdf

NTSB. (1998). LAX98FA008. Washington, D.C.: National Transportation Safety Board. https://www.ntsb.gov/about/employment/_layouts/ntsb.aviation/brief2.aspx?ev_id=20001208X09045&ntsbno=LAX98FA008&akey=1

NTSB. (2008). Aircraft accident investigation report - Boeing 737-4Q8, PK-KKW, Makassar Strait, Sulawesi, Republic of Indonesia, 1 January 2007 (No. KNKT/07.01/08.01.36): National Transportation Safety Committee, Ministry of Transportation, Republic of Indonesia. https://www.atsb.gov.au/media/24500/Final_ AdamAir_PK_KKW_ 17032008.pdf

NTSB. (2014a). In-flight breakup during test flight scaled composites spaceshiptwo, N339SS, Near Koehn Dry Lake, California, October 31, 2014. Spacecraft accident report, NTSB/AAR-15/02. Washington, DC: NTSB. https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1502.pdf

NTSB. (2014b). Descent below visual glidepath and impact with seawall. Asiana Airlines Flight 214, Boeing 777-200ER, HL7742. San Francisco, California, July 6, 2013. Aircraft Accident Report, AAR-14/01. Washington, DC: NTSB. https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1401.pdf

Null, C. H, Kaiser, M. K., Zaal., P. M., Dempsey, D. L., Vos, G. A., Mulligan, J. B., & Billimoria, K. D. (2019). *Simulator fidelity requirements for the human landing system (HLS)* [Unpublished report]. NASA NESC human factors technical discipline team.

Nunes, A. (2003, October). *The impact of automation use on the mental model: Findings from the air traffic control domain*. Presented at the proceedings of the human factors and ergonomics society 47th annual meeting.

O'Connor, B. (2006, May). *A gift: Lessons and analogies from STS-3 close call*. NASA. https://sma.nasa.gov/safety-messages/a-gift.

Oberg, J. (2004). Space station crew perpares to fix window. In: NBCNEWS.COM. http://www.msnbc.msn.com/id/4253674/ns/technology_and_science-space/t/space-station-crew-prepares-fix-window/ Accessed February 12 .

Oerbeck, B., Sundet, K., Kase, B. F., & Heyerdahl, S. (2003). Congenital hypothyroidism: influence of disease severity and L-thyroxine treatment on intellectual, motor, and school-associated outcomes in young adults. *Pediatrics, 112*(4), 923-930. https://doi.org/10.1542/peds.112.4.923

Oglesby, J. M., Stowers, K., Leyva, K., Dietz, A., Sonesh, S., Burke, S., & Salas, E. (2014). Assessing human-automation system safety, efficiency, and performance: Developing a metrics framework. *Proceedings of the human factors and ergonomics society annual meeting*, *58*(1), 1149-1153. https://doi.org/10.1177/1541931214581240

Oleson, K. E., Billings, D. R., Kocsis, V., Chen, J. Y. C., & Hancock, P. A. (2011). Antecedents of trust in human-robot collaborations. *Proceedings of 2001 IEEE international multi-disciplinary conference on cognitive methods in situation awareness and decision support (CogSIMA),* 175-178. doi: 10.1109/COGSIMA.2011.5753439.

Olsen, D. R., Jr., & Wood, S. B. (2004). Fan-out: Measuring human control of multiple robots. *Proceedings of the SIGCHI conference on human factors in computing systems*, 231-238. https://doi.org/10.1145/985692.985722

Olsen, J. & Goodrich, M. A., (2003). *Metrics for evaluating human-robot interactions*. Presented at the performance metrics for intelligent systems (PERMIS). PERMIS23.

Oman, C. & Liu, A. (2018). Design and automation of automated electronic checklists for robotic operations. [Unpublished report]. NASA Johnson Space Center.

Omidshafiei, S., Agha-Mohammadi, A., Chen, Y. F., Üre, N. K., How, J. P., Vian, J. L., & Surati, R. (2015). MAR-CPS: Measurable augmented reality for prototyping cyber-physical systems. *Proceedings of the AIAA SciTech*. *AIAA Infotech @ Aerospace*. AIAA 2015-0643. http://dx.doi.org/10.2514/6.2015-0643

Ong, K. W., Seet, G., & Sim, S. K. (2008). An implementation of seamless human-robot interaction for telerobotics. *International journal of advanced robotic systems*, *5*(2), 167-176. An implementation of seamless human-robot interaction for telerobotics. I

Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human performance consequences of stages of stages and levels of automation: An integrated meta-analysis. *Human factors*, *56*(3), 476-488. https://doi.org/10.1177/0018720813501549

Opperman, R. (1994). *Adaptive user support*. Hillsdale, NJ: Erlbaum.

Orasanu, J., Fischer, U. (1992). Distributed cognition in the cockpit: Linguistic control of shared problem solving. In *Proceedings of the Fourteenth annual conference of the cognitive science society*, pp. 189–194. Hillsdale, NJ: Erlbaum.

Orloff, R. W. (2000). *Apollo by the numbers: a statistical reference*. NASA report, NASA SP-2000-4029. Washington, DC: NASA. https://history.nasa.gov/SP-4029.pdf

Ortega, H. & Harm, D. (2008). Space and entry motion sickness. In M. R. Barratt & S. L. Pool, (Eds.), *Principles of clinical medicine for space flight*, pp. 211-222. Springer.

Ososky, S., Philips, E., Schuster, D., & Jentsch, F. (2013a). A picture is worth a thousand mental models: Evaluating human understanding of robot teammates. *Proceedings of the human factors and ergonomics society annual meeting*, *57*(1), 1298-1302. https://doi.org/10.1177/1541931213571287

Ososky, S., Schuster, D., Phillips, E., & Jentsch, F. (2013b). *Building appropriate trust in human-robot teams* [Oral presentation] the AAAI spring symposium series. https://www.aaai.org/ocs/index.php/SSS/SSS13/paper/view/5784

Palinkas, L. & Suedfeld, P. (2008). Psychological effects of polar expeditions. *The lancet (London, England), 37*1(9607), 153-163. https://doi.org/10.1016/S0140-6736(07)61056-3 t, 371, 153-163.

Panontin, T., Wu, S.C, & Vera, A. (2021). *Interim report on number and types of items found in the IFI database search*, NASA Ames Research Center, July.

Parasuraman, R., & Hancock, P. A. (2007). Mitigating the adverse effects of workload, stress, and fatigue with adaptive automation. In P.A. Hancock & J.L. Szalma (eds), P*erformance under stress*, pp. 45-57. Burlington, VT: Ashgate.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors, 39*(2), 230-253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Galster, S., Squire, P., Furukawa, H., & Miller, C. (2005). A flexible delegation-type interface enhances system performance in human supervision of multiple robots: Empirical Studies with RoboFlag. *IEEE transactions on systems, man, and cybernetics - part a: systems and humans*, *35*(4), 481-493. https://doi.org/10.1109/TSMCA.2005.850598

Parasuraman, R. (2011). Neuroergonomics: Brain, cognition, and performance at work. *Current directions in psychological science*, *20*(3), 181-186. https://doi.org/10.1177/0963721411409176

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). 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), 286-297. DOI: 10.1109/3468.844354

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced complacency. *International journal of aviation psychology, 3*(1), 1-23. https://doi.org/10.1207/s15327108ijap0301_1

Parasuraman, R., & Wilson, G. F. (2008). Putting the brain to work: Neuroergonomics past, present, and future. *Human factors*, *50*(3), 468-474. https://doi.org/10.1518/001872008X288349

Park, S. H., & Woldstad, J. C. (2000). Multiple two-dimensional displays as an alternative to three-dimensional displays in telerobotic tasks. *Human factors*, *42*, 592-603. https://doi.org/10.1518/001872000779698060

Patrashkova-Volzdoska, , R.R., McComb, S. A. , Green, S. G., & Compton, W. D. (2003). Examining a curvilinear relationship between communication frequency and team performance in cross-functional project teams. In IEEE transactions on engineering management, vol. 50(3), pp. 262-269. doi: 10.1109/TEM.2003.817298.

Pazuchanics, S. L. (2006). The effects of camera perspective and field of view on performance in teleoperated navigation. *Proceedings of the human factors and ergonomics society's 50th annual meeting*, *50*(16), 1528-1532. https://doi.org/10.1177/154193120605001603

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations*, pp. 47-87. Cambridge, UK: Cambridge University Press.

Peck, E. M., Afergan, D., Yuksel, B. F., Lalooses, F., & Jacob, R. J. (2014). Using fNIRS to measure mental workload in the real world. In S. H. Fairclough & K. Gilleade (eds.), *Advances in physiological computing*, pp. 117-139. Springer, London.

Pederson, L., Kortencamp, D., Wettergreen, D., & Nourbakhsh, I. (2003). A survey of space robotics. *Proceedings of the 7th International symposium on artificial intelligence, robotics, and automation in space*. http://wayback.archive-it.org/1792/20100207195709/http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20030054507_2003061124.pdf

Phillips, E., Ososky, S., Grove, J., & Jentsch, F. (2011). From tools to teammates: Toward the development of appropriate mental models for intelligent robots. *Proceedings of the human factors and ergonomics society annual meeting*, *55*(1), 1491-1495. https://doi.org/10.1177/1071181311551310.

Pijnenburg, B., & Van Duin, M. J. (1991). The Zeebrugge ferry disaster. Elements of a communication and information processes scenario. In U. Rosenthal & B. Pijnenburg (Eds.), *Crisis management and decision making: simulation oriented scenarios*, pp. 45. Springer.

Plice, L., Dono, A., & West, S. (2019, August). *HelioSwarm: Swarm Mission Design in High Altitude Orbit for Heliophysics*. NASA technical report Document ID 20190029109. https://ntrs.nasa.gov/search.jsp?R=20190029109

Prewett, M. S., Johnson, R. C., Saboe, K. N., Elliott, L. R., & Coovert, M. D. (2010). Managing workload in human-robot interaction: A review of empirical studies. *Computers in human behavior*, *26*(5), 840-856. https://doi.org/10.1016/j.chb.2010.03.010

Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2014a). Measuring human-automation function allocation. *Journal of cognitive engineering and decision making*, *8*(1), 52-77. https://doi.org/10.1177/1555343413490166

Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2014b). Modeling human-automation function allocation. *Journal of cognitive engineering and decision making*, *8*(1), 33-51. https://doi.org/10.1177/1555343413490944

Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D., Chu, H., Cosgrove, S., Sexton, B., Hyzy, R., Welsh, R., Roth, G., Bander, J., Kepros, J., & Goeschel, C. (2006). An intervention to decrease catheter-related bloodstream infections in the ICU. *The New England journal of medicine, 355*(26), 2725-2732. https://doi.org/ 10.1056/ NEJMoa061115

Proteau, L., Marteniuk, R. G., & Lévesque, L. (1992). A sensorimotor basis for motor learning: Evidence indicating specificity of practice. *Quarterly journal of experimental psychology, 44*(3), 557-575. https://doi.org/10.1080/14640749208401298

Rajulu, S., Young, K., Kim, H., & Vu, L. (2018). *Quantification of inflight physical changes: anthropometry and neutral body posture* [Unpublished report]. NASA Human research program (HRP).

Rando, C., Baggerman S. D., & Duvall L. E. (2005) Habitability in space. P*roceedings of the Human factors and ergonomics society 49th annual meeting*, *49*(1), 5-9. https://doi.org/10.1177/154193120504900103

Reason, J. (1997*). Managing the risks of organizational accidents*. Aldershot: Ashgate.

Rickard, T. C., & Bourne, L. E., Jr. (1995). An identical-elements model of basic arithmetic skills. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity*, pp. 255-281. Thousand Oaks, CA: Sage.

Rickard, T. C., & Bourne, L. E., Jr. (1996). Some tests of an identical elements model of basic arithmetic skills. *Journal of experimental psychology: Learning, memory, and cognition, 22*, 1281-1295. https://psycnet.apa.org/doi/10.1037/0278-7393.22.5.1281

Riley, V. (1989). A general model of mixed-initiative human- machine systems. In *Proceedings of the human factors society 33rd annual meeting,* pp. 124-128. Santa Monica, CA: Human Factors and Ergonomics Society.

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger, III, & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving*, pp. 3-41. Hillsdale, NJ: Erlbaum.

Roscoe, A.H. 1984. *Assessing pilot workload in flight.* Flight test techniques, AGARD, Paris. (AGARD-CP-373).

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use*. Technical report No. TR 90019. Royal Aerospace Establishment. https://books.google.com/books?id=0_eJXwAACAAJ

Roth, E. M., & Pritchett, A. R. (2017). Preface to the special issue on advancing models of human-automation interaction. *Journal of cognitive engineering and decision making*, *12*(1), 3-6. https://doi.org/10.1177/1555343417749192

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological R\review, 103*(4), 734–760.

Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., Mutlu, B., & McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. *Computer graphics forum*, *34*(6), 299-326. https://doi.org/10.1111/cgf.12603

Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of experimental psychology: General,* 112(4), 585-615. https://doi.org/10.1037/0096-3445.112.4.585

Rushton, S. K., & Wann, J. P. (1999). Weighted combination of size and disparity: a computational model for timing a ball catch. *Nature neuroscience, 2*(2), 186-190. doi:10.1038/5750

S&MA Flight Safety Office. (2012, Spring). *Significant incidents and close calls in human spaceflight.* NASA. http://spaceflight.nasa.gov/outreach/Significant_Incidents.pdf

Salas, E., Wilson, K. A., Priest, H. A., & Guthrie J. W. (2006) Design, delivery, and evaluation of training systems. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics, Third edition*, pp. 472-512. Hoboken, N.J., John Wiley & Sons, Inc. https://doi.org/10.1002/0470048204.ch18.

Salmon, P., Stanton, N., Walker, G., & Green, D. (2006). Situation awareness measurement: a review of applicability for C4i environments. *Applied ergonomics, 37*(2), 225-238. doi: 10.1016/j.apergo.2005.02.001..

Salomon, G. (1993). No distribution without individuals' cognition: A dynamic interactional view. In G. Salomon (Ed.), *Distributed cognitions: psychological and educational considerations*, pp. 111-138. Cambridge, UK: Cambridge University Press.

Sandal, G. M. (2018). Psychological resilience. In T. Sgobba, B. Kanki, J.-F. Clervoy, & G. M. Sandal (Eds.), *Space safety and human performance*, pp. 183-237. https://doi.org/10.1016/B978-0-08-101869-9.00006-6.

Sanders, M. S., & McCormick, E. J. (1993). *Human Factors in engineering and design, 7th ed*. McGraw-Hill Science/Engineering/Math

Sándor, A., Begault, D. R., & Holden, K. (2010a). *Speech alarm study* [Unpublished report]. NASA, Johnson Space Center.

Sándor, A. & Holden, K., (2009). *Measuring legibility* [Unpublished report]. NASA, Johnson Space Center.

Sándor, A., Holden, K., & Archer, R. (2010b). Usability task. [Unpublished report]. NASA, Johnson Space Center.

Sándor, A., Holden, K., Thompson, S., Pace, J., Adelstein, B., Beutter, B., McCann, R., & Anderson, M. (2010c). *Performance with continuous and discrete cursor control device under vibration frequencies and amplitudes* [Unpublished report]. NASA, Johnson Space Center.

Sándor, A., Moses, H., Sprufera, J., & Begault, D. (2016). *Memo on speech alarms: Replication and validation of results.* NASA technical memorandum 2016-219107. https://ntrs.nasa.gov/citations/20190002684

Sandoval, L., Keeton, K., Shea, C., Otto, C., Patterson, H., & Leveton, L. (2012). *Perspectives on asthenia in astronauts and cosmonauts: Review of the international research literature*. NASA technical memorandum NASA/TM-2012-217354. https://ntrs.nasa.gov/citations/20110023297

Saner, L. D., Bolstad, C. A., Gonzalez, C., & Cuevas, H. M. (2009). Measuring and predicting shared situation awareness in teams. *Journal of cognitive engineering and decision making*, *3*(3), 280-308. https://doi.org/10.1518/155534309X474497

Sanli, E. A., & Carnahan, H. (2018). Long-term retention of skills in multi-day training contexts: A review of the literature. *International journal of industrial ergonomics*, 66, 10-17. https://doi.org/10.1016/j.ergon.2018.02.001

Sanli, E., Ennis, K. A., Brown, R., & Carnahan, H. (2018) The development and validation of key performance indicators for two tasks related to Marine Emergency Duties training, *International journal of training research, 16*(2), 145-154. https://doi.org/10.1080/14480220.2018.1501911

Sarter, N. B. & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. *Human factors, 39*(4), 553-569. https://doi.org/10.1518/001872097778667997

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human factors, 37*(1), 5-19. https://doi.org/10.1518/001872095779049516

Sarter, N. B., Woods, D. D., and Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.) *Handbook of human factors and ergonomics*, 2nd ed., pp. 1926-1943. New York: Wiley.

Sarter, N., & Woods, D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The international journal of aviation psychology*, *4*(1), 1-28. https://doi.org/10.1207/ s15327108ijap0401_1

Sauro, J., & Lewis, J. (2005) Estimating Completion Rates from Small Samples using Binomial Confidence Intervals: Comparisons and Recommendations. *Proceedings of the human factors and ergonomics society annual meeting*, Orlando, FL. https://doi.org/10.1016/j.apergo.2012.05.008

Saulnier, P., Sharlin, E., & Greenberg, S. (2011). Exploring minimal nonverbal interruption in HRI. *Proceedings f theo EEE international workshop on robot and human communication (ROMAN),* pp. 79-86. https://doi.org/10.1109/ROMAN.2011.6005257

Schaafstal, A., Schraagen, J. M., & Berlo, M. van (2000). Cognitive task analysis and innovation of training: The case of structured troubleshooting. *Human factors, 42*(1), 75-86. https://doi.org/10.1518/001872000779656570

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors*, *58*(3), 377-400. https://doi.org/10.1177/0018720816634228.

Scheuring, R. A., Mathers, C. H., Jones, J.A., & Wear, M. L. (2009). Musculoskeletal injuries and minor trauma in space: incidence and injury mechanisms in US astronauts. *Aviation, space, and environmental medicine, 80*(2):117-124. https://doi.org/10.3357/asem.2270.2009

Schmaus, P., Leidner, D., Bayer, R., Pleintinger, B., Krüger, T., & Lii, N. Y. (2019). Continued advances in supervised autonomy user interface design for METERON SUPVIS Justin. *Proceedings of the 2019 IEEE aerospace conference*, pp. 1-11. doi:10.1109/AERO. 2019.8741885

Scholtz, J. (2003). Theory and evaluation of human robot interactions. *Proceedings of the 36th Annual Hawaii international conference on system sciences*. https://doi.org/ 10.1109/HICSS.2003.1174284

Scholtz, J., Antonishek, B., & Young, J. (2005). Implementation of a situation awareness assessment tool for evaluation of human-robot interfaces. *IEEE transactions on systems, man, and cybernetics - part A: Systems and humans, 35*, 450-459.

Scholtz, J., Young, J., Drury, J. L., & Yanco, H. A. (2004). Evaluation of human-robot interaction awareness in search and rescue. *Proceedings the IEEE international conference on robotics and automation, ICRA '04. 2004*, *3*, 2327-2332. https://doi.org/10.1109/ROBOT.2004.1307409

Schroeder, J. A. & Grant, P. R.(2010). Pilot behavioral observations in motion flight simulation. *Proceedings of the AIAA guidance, naviga- tion, and control conference, Toronto, Canada, Aug. 2-5*, No. AIAA-2010-8353.

Schroeder, J. E., & Tuttle, M. L. (1992). *Investigation of possible causes for human-performance degradation during microgravity flight.* NASA technical report No. 19920012102. https://ntrs.nasa.gov/search.jsp?R=19920012102

Sebok, A., Wickens, C. D., Sarter, N., Li, H., Gore, B., Hooey, B., Gacy, M., & Brehon, M. (2013). *MIDAS-FAST: Development and validation of a tool to support function allocation* [Unpublished report]. NASA human research program and the national space biomedical research institute.

Semjen, A., Leone, G., & Lipshits, M. (1998). Temporal control and motor control: Two functional modules which may be influenced differently under microgravity. *Human movement science*, *17*(1), 77-93. https://doi.org/10.1016/S0167-9457(97)00025-0

Seo, D. W., & Lee, J. Y. (2013). Direct hand touchable interactions in augmented reality environments for natural and intuitive user experiences. *Expert systems with applications, 40*(9), 3784-3793. https://doi.org/10.1016/j.eswa.2012.12.091

Sgobba, T., & Kanki, B. (2018). Introduction. In T. Sgobba, B. Kanki, J.-F. Clervoy, & G. M. Sandal (Eds.), *Space safety and human performance*, pp. 1-16. https://doi.org/10.1016/B978-0-08-101869-9.00001-7

Shayler, D. (2000). *Disasters and accidents in manned spaceflight*. Library of Congress. http://www.loc.gov/catdir/enhancements/fy0816/99045244-d.html

Sheridan, T. (Ed.). (2002). *Human and automation: Systems design and research issues.* New York: Wiley.

Sheridan, T. B. (2016). Human-robot interaction: Status and challenges. *Human factors*, *58*(4), 525-532. https://doi.org/10.1177/0018720816644364

Sheridan, T. B., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of human factors and ergonomics*, *1*(41), 89-129. https://doi.org/10.1518/155723405783703082

Simon, M., Whitmire, A., Otto, C., & Neubek, D. (2011). Factors impacting habitable volume requirements: Results from the 2011 habitable volume workshop. NASA technical memorandum, NASA/TM-2011-217352. https://ntrs.nasa.gov/citations/20110023287

Singh, R., Miller, T., Newn, J., Sonenberg, L., Velloso, E., & Vetere, F. (2018). Combining planning with gaze for online human intention recognition. *Proceedings of the 17th international conference on autonomous agents and multiagent systems*, 488-496. ACM Digital Library. http://dl.acm.org/citation.cfm?id=3237383.3237457

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Siu, K., Best, B., Kin, J., Oleynikov, D., & Ritter, F. (2016). Adaptive virtual reality training to optimize military medical skills acquisition and retention. *Military medicine, 181*(5), 214-220. https://doi.org/10.7205/MILMED-D-15-00164

Skitka, L. J., Mosier, K., & Burdick, M. D. (1999). Does automation bias decision-making? *International journal of human-computer studies*, *51*(5), 991-1006. https://doi.org/10.1006/ijhc.1999.0252

Slack, K. J., Williams, T. J., Schneiderman, J. S., Whitmire, A. M., & Picano, J. J. (2016). *Evidence report: Risk of adverse cognitive or behavioral conditions and psychiatric disorders*. NASA. https://humanresearchroadmap.nasa.gov/Evidence/reports/BMed.pdf

Smith, R. L., & Stuart, M. A. (1989). The effects of spatially displaced visual feedback on remote manipulator performance. *Proceedings of the human factors and ergonomics society annual meeting*, *33*(20), 1430-1434. https://doi.org/10.1177/154193128903302015

Smith, C. D., Umberger, G. H., Manning, E. L., Slevin, J. T., Wekstein, D. R., Schmitt, F. A., Markesbery, W. R., Zhang, Z., Gerhardt, G. A., Kryscio, R. J., & Gash, D. M. (1999). Critical decline in fine motor hand movements in human aging. *Neurology, 53*(7), 1458-1458. https://doi.org/10.1212/wnl.53.7.1458

Smyth, C. C., Gombash, J. W., & Burcham, P. M. (2001). Indirect vision driving with fixed flat panel displays for near-unity, wide, and extended fields of camera view. Army research laboratory technical report, ARL-TR-2511. https://apps.dtic.mil/dtic/tr/fulltext/u2/a392445.pdf

Snow, M. P., & Reising, J. M. (2000, July). Comparison of two situation awareness metrics: SAGAT and SA-SWORD. *Proceedings of the human factors and ergonomics society annual meeting*, *44*(13), 49-52. https://doi.org/10.1177/154193120004401313

Sogame, H., & Ladkin, P. (1996). Aircraft accident investigation report 96-5. Ministry of Transport. http://sunnyday.mit.edu/accidents/nag-4-7.html

Society of Automotive Engineers (SAE) International (2019). *Systems Management Standard: Standard practice for human systems integration*, SAE 6906. 2019-02

Sonoda, M. (2016). *NASA's Human Spaceflight Training Reduction Effort*. [Unpublished Report]. NASA.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction. *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction*, 33-40. https://doi.org/10.1145/1121241.1121249

Sternberg, S. (1966). High-speed scanning in human memory. *Science (New York, N.Y.), 153*(3736), 652-654. https://doi.org/10.1126/science.153.3736.652

Stoll, E., Jackel, S., Katz, J., Saenz-Otero, A., & Varatharjoo, R. (2012). SPHERES interact-human-machine interaction aboard the International Space Station. *Journal of field robotics*, *29*(4), 554-575. https://doi.org/10.1002/rob.21419

Stowers, K., Oglesby, J., Sonesh, S., Leyva, K., Iwig, C., & Salas, E. (2017). A framework to guide the assessment of human-machine systems. *Human factors*, *59*(2), 172-188. https://doi.org/10.1177/0018720817695077

Strangman, G. E., Sipes, W., & Beven, G. (2014). Human cognitive performance in spaceflight and analogue environments. *Aviation, space, and environmental medicine*, *85*(10), 1033-1048. https://doi.org/10.3357/ASEM.3961.2014

Strauch, B. (1997). Automation and decision making—Lessons from the Cali accident. *Proceedings of the human factors and ergonomics 41st annual meeting, 41*(1), 195-199. *https://doi.org/10.1177/107118139704100145*

Strauch, B. (2017). The automation-by-expertise-by-training interaction: Why automation-related accidents continue to occur in sociotechnical systems. *Human factors,* 59(2), 204-208. https://doi.org/10.1177/0018720816665459.

Strauss, S. (2004). Extravehicular mobility unit training suit symptom study report. NASA technical report, NASA/TP-2004-212075. https://ntrs.nasa.gov/citations/20060017061

Stuster, J. (1996). *Bold Endeavors: Lessons from polar and space exploration*. Naval Institute Press, Annapolis, Maryland.

Stuster, J. (2010). Acceptable Risk: The human mission to Mars. *Journal of cosmology, 12*, 3566-3577. http://journalofcosmology.com/Mars122.html

Stuster, J. W., Adolf, J. A., Byrne, V. E., & Greene, M. (2018). Human exploration of Mars: Preliminary lists of crew tasks. NASA contractor report, NASA/CR-2018-220043 https://ntrs.nasa.gov/citations/20190001401

Stuster, J. W., Adolf, J. A., Byrne, V. E., & Greene, M. (2019). Generalizable skills and knowledge for exploration missions. NASA/CR-2018-220445.

Styf, J. R., Ballard, R. E., Fechner, K., Watenpaugh, D. E., Kahan, N. J., & Hargens, A. R. (1997). Height increase, neuromuscular function, and back paind during 6 degree head-down tilt with traction. *Aviation, space, and environmental medicine, 68*(1):24-29.

Sullenberger, C. B. (2010). Preparing for excellence: Leading your team to success. Proceedings of the human factors and ergonomics society 54th annual meeting. San Francisco. https://hfes.connectedcommunity.org/hfes/events/future-and-past-hfes-meetings.

Svitak, A. (2013, February 25). *HTV-3 abort caused by friction with station arm.* Retrieved April 3, 2015, from aviation week network website: http://aviationweek.com/awin/htv-3-abort-caused-friction-station-arm

Szafir, D., Mutlu, B., & Fong, T. (2015). Communicating directionality in flying robots. *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, 19-26. https://doi.org/10.1145/2696454.2696475

Takayama, L., & Pantofaru, C. (2009). Influences on proxemic behaviors in human-robot interaction. *Proceedings of the international conference on intelligent robots and systems - IROS*, 5495-5502. https://doi.org/10.1109/IROS.2009.5354145.

Taylor, R. M. (1990). Situation awareness rating technique (SART): the development of a tool for aircrew systems design. In *Situational awareness in aerospace operations*, chapter 3. France: Neuilly sur-Seine, NATO-AGARD-CP-478.

Tayyari, F., & Smith, J.L. (1997). *Occupational ergonomics: Principles and applications*. Kluwer Academic Publishers, Boston.

Terwiesch, P. G., & Ganz, C. (2009). Trends in automation. In S. Nof (Ed.), *Springer handbook of automation*, pp. 127-143. Berlin: Springer-Verlag.

Thapan, K., Arendt, J., & Skene, D. (2010). An action spectrum for melatonin suppression: evidence for a novel non-rod, non-cone photoreceptor system in humans. *Journal of physiology 535*, 261-267. https://doi.org/10.1111/j.1469-7793.2001.t01-1-00261.x

Thaxton, S., & Chen, M. (2013). 2012 Habitable volume workshop summary report. Houston, TX. http://houstonhfes.org/conferences/ conference2013/Proceedings/ HHFES%202013%20HV%20Workshop%20Thaxton.pdf

Thaxton, S., Chen, M., Hsiang, S., Lim, C., Meyers, J. and Wald, S., (2017). Spacecraft Optimization Layout and Volume (SOLV): Development of a model to assess habitable volume. *Proceedings of the 2017 IEEE aerospace conference*, Big Sky, MT, 2017, pp. 1-16. Doi: 10.1109/AERO.2017.7943592.

Thaxton, S., Chen, M., & Whitmore, M. (2014). *2012 Habitable volume workshop results: Technical products prepared for human research program (HRP) space human factors engineering (SHFE) portfolio.* https://www.semanticscholar.org/paper/2012-Habitable-Volume-Workshop-Results%3A-Technical-Thaxton-Chen/aa4d6cac590cc15fbd76eeca49f536d292f96211

Thaxton, S. S., Litaker, H .L., Holden, K. L., Adolf, J. A., & Morency, R. M. (2012, February). *Developmental testing of habitability and human factors tools and methods during NEEMO15.* [Oral presentation] NASA human research program investigators' workshop, Houston, TX, 2012.

Thomas, E. A., & Weaver, B. (1975). Cognitive processing and time perception. *Perception & psychophysics, 17*(4), 363-367. https://doi.org/10.3758/BF03199347

Thomas, L. C., & Wickens, C. D. (2000). Effects of display frames of reference on spatial judgments and change direction. Army research laboratory technical report, ARL-00-14/FED-LAB-00-4. https://apps.dtic.mil/dtic/tr/fulltext/u2/a436771.pdf

Thompson, S., Holden, K., & Sándor, A. (2015). *Validation of the fine motor skills application using the 9-hole peg test* [Unpublished report]. NASA, Johnson Space Center.

Thompson, S., Holden, K., Ebert, D., Root, P., Adelstein, B., & Jones, J. (2010). *Short duration bioastronautics investigation (SDBI) 1904: Human factors assessment of vibration effects on visual performance during launch.* [Poster presentation] 2010 human factors & ergonomics society houston chapter meeting, Houston, TX.

Thorndike, E. L. (1906). *The principles of teaching: Based on psychology*. New York: A. G. Seiler.

Thornton, W., Hoffler, G., & Rummel, J. (1977). Anthropometric changes and fluid shifts. In: Johnston R, Dietlin L (Eds.), *Biomedical results from Skylab*. NASA, Washington, D.C., pp. 330-338.

Thornton, W., & Moore, T. (1987). Height changes in microgravity. In M.W. Bungo & M. A. Bowman (Eds.), *Results of the life sciences DSO's conducted aboard the Space Shuttle 1981-1986*, pp. 55-57. NASA, Washington, D.C.

Tittle, J. S., Roesler, A., & Woods, D. D. (2002). The remote perception problem. *Proceedings of the human factors and ergonomics society annual meeting*, *46*(3), 260-264. https://doi.org/10.1177/154193120204600309

Trafton, G., Cassimatis, N., Bugajska, M., Brock, D., Derek, P., Mintz, F., & Schultz, A. (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE transactions on systems man and cybernetics - part a systems and humans, 35*(4), 460-470. DOI: 10.1109/TSMCA.2005.850592

Transportation Safety Board of Canada. (2008). *Encounter with wake turbulence, Air Canada, Airbus A319-114 C-GBHZ.* Aviation investigation report no. A08W0007. Washington State, United States. https://www.tsb.gc.ca/eng/rapports-reports/aviation/2008/a08w0007/a08w0007.html

Tucker, A. M., Whitney, P., Belenky, G., Hinson, J. M., & Van Dongen, H. P. (2010). Effects of sleep deprivation on dissociated components of executive functioning. *Sleep*, *33*(1), 47-57. https://doi.org/10.1093/sleep/33.1.47

Tullis, T., & Albert, W. (2008). Measuring the user experience: Collecting, analyzing, and presenting usability metrics. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological review, 80*(5), 352-373. https://doi.org/10.1037/h0020071

U.S. Department of Energy (2013) *Building technologies office solid-state technology fact sheet*. PNNL-22908**.** Pacific Northwest National Laboratory.

Urbina, M. & Fleming, N. (2019) *Medial system concept of operations for mars exploration mission-11*. NASA, Houston, TX, JSC 67070.

Velazquez, J., Peck, A. G., & Sestak, T. (2015). Behavioral traps in crew-related aviation accidents. *Proceedings of the 18th international symposium on aviation psychology,* pp. 372-377. Dayton, OH: Wright State University.

Vanoutryve, B., Rosa, D. D., Fisackerly, R., Houdou, B., Carpenter, J., Philippe, C., Pradier, A., et al. (2010). An analysis of illumination and communication conditions near lunar south pole based on Kaguya data. *Proceedings of the international planetary probe workshop 2010 (IPPW-7)*, pp. 7.

Vera, A. H., Holden, K. L., Dempsey, D. L., Russi-Vigoya, M. N., Wu, S. & Beutter, B. R. (2019). *Contextual inquiries and interviews to support crew autonomous operations in future deep space missions: Preliminary requirements and proposed future research.* [Unpublished reprt]. NASA.

Vera. A. H. (2019) *Risk of adverse outcome due to inadequate human systems integration (HSI) architecture* [Unpublished report]. NASA human system risk board.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Library of Congress. http://www.loc.gov/catdir/ enhancements/ fy0740/98031670-d.html

Vogt, H. C., Krause, H. E., Hohlweck, H., & May, E. (1973). Mechanical impedance of supine humans under sustained acceleration. *Aerospace medicine 44*(2), 675-679.

Vogel, R. J., & Thompson, C. P. (1995). The specificity and durability of Rajan's memory. In A.F. Healy, & L. E. Bourne, Jr. (Eds). *Learning and memory of knowledge and skills: Durability and specificity*, pp. 320-342. Thousand Oaks, CA: Sage Publications.

Voshell, M., Woods, D. D., & Phillips, F. (2005). Overcoming the keyhole in human-robot coordination: Simulation and Evaluation. *Proceedings of the human factors and ergonomics society annual meeting*, *49*(3), 442–446. https://doi.org/10.1177/154193120504900348.

Vykukal, H. C. (1968). Dynamic response of the human body to combined vibration when combined with various magnitudes of linear acceleration. *Aerospace medicine, 39*(11), 1163-1166.

Wang, J., Lewis, M., & Hughes, S. (2004). Gravity-referenced attitude display for teleoperation of mobile robots. *Proceedings of the human factors and ergonomics society annual meeting*, *48*(23), 2662-2666. https://doi.org/10.1177/154193120404802312.

Watson, T. (2007). Size-conscious NASA will need astronauts who fit the profile USA Today. p 1. Baltimore, MD.

Weaver, E. (Director). (2015). *Apollo Astronauts: Training NASA's Moon Men* [Film]. United States: Jansen Media.

Webb, S., Whitley, J., & Winters, A. (2009). *Generic ground rules & relevant information: End to end maintenance scenario development* [PowerPoint slides]. NASA. https://iss-www.jsc.nasa.gov/nwo/seio/extcarr/bbt_docs/agen/pres/winters081308.ppt

Weiss, A., Bernhaupt, R., Lankes, M., & Tscheligi, M. (2009). The usus evaluation framework for human-robot interaction. *In AISB2009: Proceedings of the symposium on new frontiers in human-robot interaction*. https://www.researchgate.net/publication/313559458_The_USUS_evaluation_framework_for_human-robot_interaction.

Wenzel, E. M. (2019). *Human factors standard measure recommendations for spaceflight, analogs, and other research studies* [Unpublished report]. NASA.

Weybrew, B.B. (1979). *History of Military Psychology at the U.S. naval ubmarine medical research laboratory*. Groton, CT.

Whitmire, A., Leveton, L., Broughton, H., Basner, M., & Kearney, A. (2015). *Minimum acceptable net habitable volume for long-duration exploration missions*. NASA document number 2014001695a. https://ntrs.nasa.gov/citations/20140016951

Whitney, P., Hinson, J. M., Jackson, M. L., & Van Dongen, H. P. (2014). Feedback blunting: Total sleep deprivation impairs decision making that requires updating based on feedback. *Sleep*, *38*(5), 745-754. https://doi.org/10.5665/sleep.4668

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & cognition, 2*(4), 775-780. https://doi.org/10.3758/BF03198154

Wickens, C. D. (1991). Processing resources and attention. In: D. L. Damos (Ed*.) Multiple-task performance*, pp. 3-34. Bristol, PA: Taylor & Francis.

Wickens, C. D. (1992). Process control and automation. In C. D. Wickens (Ed.), *Engineering psychology and human performance*, 2nd ed., pp. 504-551. New York, NY: Harper Collins.

Wickens, C. D. (2005). Attentional tunneling and task management. *Proceedings of the 2005 International Symposium on Aviation Psychology*, 812-817. https://corescholar.libraries.wright.edu/isap_2005/121

Wickens, C. D., & Dixon, S. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical issues in ergonomics science*, *8*, 201-212. https://doi.org/10.1080/14639220500370105

Wickens, C. D., Li, H., Santamaria, A., Sebok, A., & Sarter, N. B. (2010). *Stages and levels of automation: An Integrated meta-analysis*.[Oral presentation]. Human factors and ergonomics society 54th annual meeting, San Francisco, CA.

Wickens, C.D., Gutzwiller, R., & Santamaria, A. (2015a). Discrete task switching in overload: A meta-analyses and a model. *International journal of human-computer studies*, *In Press*. http://dx.doi.org/10.1016/j.ijhcs.2015.01.002

Wickens, C. D., Gutzwiller, R. S., Vieane, A., Clegg, B. A., Sebok, A., & Janes, J. (2016). Time sharing between robotics and process control: Validating a model of attention switching. *Human factors*, *58*(2). https://doi.org/10.1177/0018720815622761

Wickens, C. D., Hutchins, S., Laux, L., & Sebok, A. (2015b). The impact of sleep disruption on complex cognitive tasks: A meta-analysis. *Human factors*, *57*(6), 930-946 https://doi.org/10.1177%2F0018720815571935

Wickens, C.D., Laux, L., Hutchins, S., & Sebok, A. (2014a). Effects of sleep restriction, sleep inertia, and overload on complex cognitive performance before and after workload transition: A meta analysis and two models. [Oral presentation]. *Proceedings of the human factors and ergonomics society annual meeting.*

Wickens, C.D., Li, H., Sarter, N., Sebok, A., & Gacy, M. (2014b). Validation of Scanning and Complacency Models for Robotic Arm Control. [Oral presentation]. *Proceedings of the human factors and ergonomics society annual meeting.*

Wickens, C.D., Sebok, A., Li, H., Sarter, N., & Gacy, A. (2015c). Using Modeling and Simulation to Predict Operator Performance and Automation-Induced Complacency with Robotic Automation: A Case Study and Empirical Validation. *Human factors*, 57(6), 959–975. https://doi.org/10.1177/0018720814566454

Wickens, C.D., Vieanne, A., Clegg, B., Sebok, A., & Janes, J. (2015d). Visual attention allocation between robotic arm and environmental process control: Validating the STOM task switching model. *Proceedings of the human factors and ergonomics society annual meeting*, *59*(1), 617-621. https://doi.org/10.1177/1541931215591135

Wiese, E., Wykowska, A., & Müller, H. J. (2013). Making eyes with robots: Readiness to engage in Human-Robot-Interaction depends on the attribution of intentionality. *Proceedings of the human factors and ergonomics society annual meeting*, *57*(1), 1174-1178. https://doi.org/10.1177/1541931213571261.

Williams, D., & Johnson, B. (2003). *EMU shoulder injury tiger team report.* NASA technical memorandum, NASA/TM-2003.212058. https://ntrs.nasa.gov/citations/20040200968

Wilson, W. C., Coffey, N. C., & Madaras, E. I. (2008). Detection and location technology assessment for aerospace applications leak. NASA technical memorandum, NASA/TM-2008-215347. https://ntrs.nasa.gov/citations/20080041598

Winfred, A., Jr., Winston, B., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human performance, 11,* 57-101. https://doi.org/10.1207/s15327043hup1101_3

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological science, 18*(2), 133-134. https://doi.org/10.1111/j.1467-9280.2007.01862.x

Wolf, R. (2017, November/December). *Squad overmatch enhancing 25ID team performance* [Oral presentation]. I/ITSEC 2017, Orlando, FL.

Woodling, C. H., Faber, S., Van Bockel, J. J., Olasky, C. C., Williams, W. K., Mire, J. L. C., & Homer, J. R. (1973). *Apollo experience report – Simulation of manned space flight for crew training.* NASA technical note, NASA-TN-D-7112. https://www.hq.nasa.gov/alsj/NASATND7112.pdf

Woods, D. D. (2016). The risks of autonomy: Doyle's Catch. *Journal of cognitive engineering and decision making*, *10*(2), 131-133. https://doi.org/10.1177/1555343416653562

Woods, D. D., & Dekker, S. (2000). Anticipating the effects of technological change: A new era of dynamics for human factors. *Theoretical issues in ergonomics science*, *1*(3), 272-282. https://doi.org/10.1080/14639220110037452

Woods, D., Dekker, S., Cook, R., Johannesen, L., & Sarter, N. (2010). *Behind Human Error* (2nd ed.). Ashgate Publishing.

Wu, S.-C., & Vera, A. H. (2019). S*upporting crew autonomy in deep space exploration: preliminary onboard capability requirements and proposed research questions*. Technical report of the autonomous crew operations technical interchange meeting. NASA technical memorandum, NASA/TM—2019–220345. https://humansystems.arc.nasa.gov/publications/NASA_TM-2019_220345.pdf

Yan, J. H., Rountree, S., Massman, P., Doody, R. S., & Li, H. (2008). Alzheimer's disease and mild cognitive impairment deteriorate fine movement control. *Journal of psychiatric research, 42*(14), 1203-1212. https://doi.org/10.1016/j.jpsychires.2008.01.006

Yanco, H. A., & Drury, J. (2004). Classifying human-robot interaction: An updated taxonomy. *2004 IEEE international conference on systems, man and cybernetics, 3*, 2841-2846. https://doi.org/10.1109/ICSMC.2004.1400763.

Yancey, W.L. (1971). Architecture, interaction and social control: The case of a large-scale public housing project. *Environment and behavior, 3*(1):3-21. https://doi.org/10.1177/001391657100300101

Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *HumanfFactors*, *4(3)3*, 355-365. https://doi.org/10.1518/001872001775898269

Young, K., & Rajulu, S. (2012, February). *The effects of microgravity on seated height (Spinal Elongation)*. [Oral presentation] 2012 NASA human research program investigators' workshop, Houston, TX.

Zaal, P., Popvic, A., & Zavala, M. (2015). Effects of False tilt cues on the training of manual roll control skills. *Proceedings of the AIAA Modeling and simulation technologies conference,* AIAA 2015-0655. https://doi.org/10.2514/6.2015-0655

Zak, A. (2009, March 25). Mission of Soyuz TMA-14. Retrieved March 30, 2015, from Russian space web website: http://www.russianspaceweb.com/iss_soyuztma14.html

Zhang, J., & D. A. Norman (1994). Representations in distributed cognitive tasks. *Cognitive science,* 18(1), 87-122. https://doi.org/10.1016/0364-0213(94)90021-3

Appendix A - List of Observations/Interviews Completed

Specialty	Location	Simulation	Contextual inquiry	Interview
Robotics Operations Systems Officer (ROBO)	NASA Johnson Space Center (JSC) - Space Station Training Facility (SSTF). ROBO backroom.	Contextual Inquiry: Generic – ISS R14 Stage HTV Install w/VV Sim #47 SSTF RTS 1 (B5S/2nd Floor),WFCR	07/06/18	
Environmental and Thermal Operating Systems (ETHOS)	NASA Johnson Space Center (JSC) - Space Station Training Facility (SSTF). Instructor side.	Generic - ISS R14 Stage HTV Install w/VV Sim #52 SSTF RTS 1 (B5S/2nd Floor)	07/20/18	
Biomedical Engineer (BME)	NASA Johnson Space Center (JSC) - Space Station Training Facility (SSTF). Instructor side.	Generic - ISS TS21 Stage JMST 274 w/ESA #10 SSTF TS21 SIM String 02,SSTF TS21 SIM String 04,WFCR	07/26/18	
Operations Support Officer (OSO)	NASA Johnson Space Center (JSC) - Space Station Training Facility (SSTF). Instructor side.	Generic - ISS TS21 Stage JMST 274 w/ESA #10 SSTF TS21 SIM String 02,SSTF TS21 SIM String 04,WFCR	07/26/18	
Extravehicular Activity Officer (EVA)	NASA Johnson Space Center (JSC) - Space Station Training Facility (SSTF). Instructor side.	Generic - ISS TS21 TBD Stage EVA Task w/MCE SSTF TS21 SIM String 04,WFCR	08/03/18	09/27/18
Spacecraft Communicator (CAPCOM)	NASA Johnson Space Center (JSC) - Space Station Training Facility (SSTF).Front room	Flight Control room CAPCOM	08/09/18	
Flight directors	NASA Johnson Space Center (JSC)	Interview only		09/27/18 12/06/18
Behavioral Health and Performance (BHP)	NASA Johnson Space Center (JSC)	Interview only		07/18/18
Visiting Vehicle Office (VVO)	NASA Johnson Space Center (JSC)	Interview only		08/01/18

Mission Evaluation Room managers	NASA Johnson Space Center (JSC)	Interview only		09/27/18
Safety Office - Crew-Related Anomalies	NASA Johnson Space Center (JSC)	Interview only		9/24/18
Marine officer	NAVSea – Phone Interview	Interview only		7/30/18
NASA's Global Hawk		Interview only?	Observation ?	

Appendix B – Questionnaire for Guided Interviews/Observations

A. FOR FLIGHT CONTROLLERS:

GOAL OF TODAY: We would like to understand how decision-making and problem resolution happens today, and your thoughts on how autonomous crew may do it in future missions.
For each failure or event that happens today, we'd like to understand:

1. Role of team members (who, hierarchy, roles)
2. What information/tools are used (what, where, procedures, software, data flows)?
3. Time table for resolution (what is time critical?)
4. Team and training issues
5. How this might be different for future autonomous crews

After a failure:

1. How do you recognize and (dis)confirm the failure?
2. How do you determine if there are any immediate crew actions required for safety?
3. How do you determine what functionality/capability has been affected? Immediate vs. near-future impacts?
4. How do you determine what the times-to-effect (your clocks) are?
5. How do you determine what the critical circumstances are?
6. How do you determine what your immediate goal is? Options and risks?
7. How do you determine what the Benefit/Cost/Risk trades are?
8. How do you determine what your contingency plan is?
9. How do you check/verify your assessment?

General:

1. Do you have a formal or informal list of the top types of failures that you deal with in ROBO?
2. What kind of training on problem solving or decision making have you had?
3. Please tell me a story about a time when you had to learn something new or implement a personal strategy to get things done?

4. What would you say works really well on the ROBO team in terms of problem resolution? Walk me through a process or strategy that works really well.
5. What could work better?
6. Imagine a future scenario where crew had to operate autonomously for a period of time. What do you think they would need to be successful?
    - Personnel/roles
    - Training
    - Capabilities/tools/technologies
    - Information resources
    - Other?

7. Keeping in mind conditions and problems encountered during missions, what are some ways that a smart environment or habitat could make things easier, more comfortable, and/or safer for crew when living in space?

Training:

Consequence

Which tasks in your discipline pose the highest risk to the crew or vehicle? Specifically, which tasks:

- ➢ Have a high- or medium-risk consequence if performed incorrectly
- ➢ Require an immediate or fast response time

Are there low-risk tasks that require an immediate or fast response time, require a high level of accuracy, are difficult to train, or are likely to have skills degrade?

Likelihood

Do you have historical data on occurrence of these events (to determine a likelihood of the occurrence of the planned task or unplanned incident that precipitated the task)? Do you have a MER or engineering point of contact who may have such information?

Do you have historical onboard performance data for these events (to determine a likelihood of performance error)? Did the ground intervene? What would the consequence have been had the ground not intervened?

Pre-Mission and In-Mission Training Requirements for DST Missions

Pre-Mission Training and Evaluation

1. What pre-mission training is provided for your high- and medium-risk tasks?
2. What are the evaluation performance measures, and what are your evaluation tools?
3. Are there existing requirements and guidelines for the training design, including evaluation measures?

Pre-Mission Training Facilities

1. What pre-mission training facilities are used to train high- and medium-risk tasks?

2. What is the fidelity of each of the training facilities, and is there documentation on each facility?
3. What are the performance measurement capabilities of these facilities (e.g., errors, time)?
4. What performance measures are missing?

<u>In-Mission Training and Evaluation</u>

1. What in-mission proficiency training is provided your high- and medium-risk tasks?
2. What are the evaluation performance measures, and what are your evaluation tools?
3. Are there existing requirements and guidelines for proficiency training design, including evaluation measures?

<u>In-Mission Training Facilities</u>

Identify the JITT tools and complete the information in each column of the JITT spreadsheet.

1. What onboard training facilities are used?
2. What is the fidelity of the onboard training facilities, and is there documentation?
3. What are the performance measurement capabilities (e.g., errors, time)?
4. What performance measures are missing?
5. Can you provide us with documentation on your JITT tool capabilities (i.e., technical specs, research papers, etc…), including the tasks trained?

B. FOR Mission Evaluation Room MANAGERS

Mission Evaluation Room manager questions

GOAL OF TODAY: We would like to understand how decision-making and problem resolution happens, and your thoughts on how autonomous crew may do it in future missions.

For each failure or event that happens, we'd like to understand:

6. Role of team members (who, hierarchy, roles)
7. What information/tools are used (what, where, procedures, software, data flows)?
8. Time table for resolution (what is time critical?)
9. Team and training issues
10. How this might be different for future autonomous crews

<u>After a failure:</u>

10. How do you recognize and (dis)confirm the failure?
11. How do you determine if there are any immediate crew actions required for safety?
12. How do you determine what functionality/capability has been affected? Immediate vs. near-future impacts?
13. How do you determine what the times-to-effect (your clocks) are?
14. How do you determine what the critical circumstances are?

15. How do you determine what your immediate goal is? Options and risks?
16. How do you determine what the Benefit/Cost/Risk trades are?
17. How do you determine what your contingency plan is?
18. How do you check/verify your assessment?
19. When you need additional information to solve a problem, who do you contact?

General:

8. Do you have a formal or informal list of the top types of failures that you deal with?
9. What kind of training on problem solving or decision making have you had?
10. Please tell me a story about a time when you had to learn something new or implement a personal strategy to get things done
11. What would you say works really well for the MER team in terms of problem resolution? Walk me through a process or strategy that works really well.
12. What could work better?
13. Imagine a future scenario where crew had to operate autonomously for a period of time. What do you think they would need to be successful?
    - Personnel/roles
    - Training
    - Capabilities/tools/technologies
    - Information resources
    - Other?
14. Keeping in mind conditions and problems encountered during missions, what are some ways that a smart environment or habitat could make things easier, more comfortable, and/or safer for crew when living in space?
15. How much time (on the ground) is spent looking for relevant data rather than understanding/analyzing that data?

C. FOR NAVSEA MARINE OFFICER

Thanks so much for talking with us!

Goal of Today: We would like to understand how decision-making and problem resolution happens today in autonomous submarine operations.

In general, for a typical failure or incident, we'd like to understand:

- Role of team members (who, hierarchy, roles)
- What types of information/tools are used (what, where, procedures, software, data flows)?
- Time table for resolution (what is time critical?)
- Team and training issues

<u>Specific Questions</u>

1. What types of failures or incidents do you typically have?

2. Can you describe the troubleshooting/problem resolution process? How do you determine what the problem is, safety implications, time criticality, benefit/cost/risk trades, potential contingency plans?

3. How does the process of problem solving change when you are running silent vs. non-silent operations?

4. How do you check/verify your assessment of the problem – what processes, individuals, or groups are involved? What is the hierarchy?

5. What tools or artifacts do you use during problem solving? Procedures? Flowcharts, Databases? Cue cards? Personal cheat sheets?

6. What kind of training on problem solving or decision making does the team have, if any?

7. Do you have an example of a time that you had to learn something new or implement a personal strategy to get things done?

8. What would you say works really well in your autonomous environment (capabilities or process)? What could work better?

9. To what extent do you have a "smart" environment (automatic sensing/intelligent systems)? Do you see the need for such systems in the future?

10. How do crew connect with the most important people in their lives during autonomous operations?

14 LIST OF ACRONYMS

AAIASB	Air Accident Investigation and Aviation Safety Board (Greece)
AAIB	Air Accidents Investigation Branch (United Kingdom)
AD	Alzheimer's Disease
AI	Artificial Intelligence
ALPA	Air Line Pilots Association
AMCL	Accepted Medical Conditions List
AR	Augmented Reality
ARC	Ames Research Center
ASCAN	Astronaut Candidates
BCI	Brain-Computer Interface
C&W	Cautions and Warnings
CAA	Civil Aviation Authority (United Kingdom)
CapCom	Capsule Communicator
CM	Countermeasure
ConOps	Concept of Operations
CPR	Cardiopulmonary Resuscitation
CSPO	Closely Spaced Parallel Operations
CCD	Cursor Control Device
CH	Congenital Hypothyroidism
CLA	Coupled Loads Analysis
COBRA	Comprehensive Oculomotor Behavioral Response Assessment
CRM	Crew Resource Management
CxP	Constellation Program
DAG	Directed Acyclic Graph
DoD	Department of Defense
DOF	Degree of Freedom
DOUG	Dynamic Ubiquitous On-Board Graphics
DOT	Department of Transportation
DRM	Design Reference Mission
DST	Deep Space Transport
DSB	Dutch Safety Board
EEG	Electroencephalograms
DTO	Detailed Technical Objective
EMU	Extra-vehicular Mobility Unit
EVA	Extra Vehicular Activity
EVR	Extravehicular Robotics
ExMC	Exploration Medical Capability
EMG	Electromyogram
FAA	Federal Aviation Administration
FAST	Future Attribute Screening Technology
FCI	Flight Crew Integration

fNIRS	Functional Near-Infrared Spectroscopy
FOD	Flight Operations Directorate
g	Gravity
GOMS	Goals, Operations, Methods, and Selection
GPWS	Ground Proximity Warning System
GUI	Graphical User Interface
HAI	Human-Automation Integration
HARI	Risk of Inadequate Design of Human and Automation / Robotic Integration
HCD	Human-Centered Design
HCAAM	Human Capabilities Assessment for Autonomous Missions
HCI	Human-Computer Interaction
HCTS	Human-Computer Trust Scale
HERA	Human Exploration Research Analog
HFACS	Human Factors Analysis and Classification System
HFBP	Human Factors and Behavioral Performance
HHC	Human Health Countermeasures
HITL	Human-in-the-loop
HLS	Human Landing System
HMD	Head Mounted Display
HPM	Human Performance Model
HRF	Human Research Facility
HRI	Human-Robotic Integration
HRP	Human Research Program
HSI	Human-System Integration
HSIA	Human-System Integration Architecture
HSID	Human-System Interaction Design
HSRB	Human-System Risk Board
HUT	Hard Upper Torso
IA	Information Architecture
ICU	Intensive Care Unit
IFI	Items for Investigation
IMG	Integrated Medical Group
iSHORT	Space Habitability Observation Reporting Tool
ISO	International Standards Organization
ISS	International Space Station
JIT	Just-In-Time
JPDO	Joint Planning and Development Office
KLM	Koninklijke Luchtvaart Maatschappij (Royal Dutch Airlines)
LDEM	Long Duration Exploration Mission
LEO	Low Earth Orbit
LED	Light-Emitting Diode
LOA	Level of Automation

LOC	Loss of Crew
LOM	Loss of Mission
MCC	Mission Control Center
MCI	Mild Cognitive Impairment
MER	Mars Exploration Rovers
MIDAS	Man-Machine Integration Design and Analysis System
MMORPG	Massively Multiplayer Online Role-Playing Game
MPTASK	Risk of Inadequate Mission, Process, and Task Design
MRT	Multiple Resource Theory
NASA	National Aeronautics and Space Administration
NextGen	Next Generation
NBL	Neutral Buoyancy Laboratory
NBP	Neutral Body Posture
NHV	Net Habitable Volume
NPR	NASA Policy Requirement
NRC	National Research Council
NTSB	National Transportation Safety Board
NTSC	National Transportation Safety Committee (Indonesia)
OBSS	Orbiter Boom Sensor System
OpsCon	Operations Concept
PDA	Personal Digital Assistant
PGSC	Payload and General Support Computers
PI	Principal Investigator
PTN	Personal Tactile Navigator
PVT	Psychomotor Vigilance Test
RAD	Robot Attention Demand
RMAT	Risk Management and Analysis Tool
ROBoT-r	Robotics On-Board Trainer
RSI	Repetitive Strain Injuries
RSP	Respiratory Support Pack
RWS	Robotic Workstation
SA	Situation Awareness
SAE	Society of Automotive Engineers
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SFRM	Spaceflight Resource Management
SDBI	Short Duration Bioastronautics Investigation
SHFE	Space Human Factors Engineering
SLEEP	Risk of Performance Errors Due to Fatigue Resulting from Sleep Loss, Circadian Desynchronization, Extended Wakefulness, and Work Overload
SME	Subject Matter Expert
SMM	Shared Mental Model

SMS	Shuttle mission simulator
SPAM	Situation Present Assessment Method
SPDM	Special Purpose Dexterous Manipulator
SPHERES	Smart Synchronized Position Hold, Engage, Reorient, Experimental Satellites
SRMS	Shuttle Remote Manipulator System
SS2	SpaceShip Two
SSRMS	Space Station Remote Manipulator System
STD	Standard
STS	Space Transportation System
TASS	Trust in Automated System Scale
TAWL	Task Analysis/Workload
TLX	Task Load Index
TRAIN	Risk of Performance Errors Due to Training Deficiencies
TSB	Transportation Safety Board
UAV	Unmanned Aerial Vehicles
VR	Virtual Reality
WHC	Waste & Hygiene Compartment