https://ntrs.nasa.gov/search.jsp?R=20170011337 2020-05-10T01:23:08+00:00Z

Autonomous Mission Operations

Jeremy Frank **NASA Ames Research Center** **Mail Stop N269-1** **Moffett Field, CA 94035-1000** **650 604 2524** **Jeremy.D.Frank@nasa.gov**	**Lilijana Spirkovska** **NASA Ames Research Center** **Mail Stop N269-1** **Moffett Field, CA 94035-1000** **650.604.4234** **lilly.spirkovska@nasa.gov**	**Rob McCann** **NASA Ames Research Center** **Mail Stop N269-1** **Moffett Field, CA 94035-1000** **650.604.0052** **Robert.s.mccann@nasa.gov**
Lui Wang **NASA Johnson Space Center** **Mail Code: ER61** **2101 NASA Parkway** **Houston, TX 77058** **lui.wang-1@nasa.gov** **281 483 8074**	**Kara Pohlkamp** **NASA Johnson Space Center** **Mail Code: DS34** **2101 NASA Parkway** **Houston, TX 77058** **kara.m.pohlkamp@nasa.gov** **281-483-2826**	**Lee Morin** **NASA Johnson Space Center** **Mail Code: CB** **2101 NASA Parkway** **Houston, TX 77058** **lee.m.morin@nasa.gov** **281 244 7970**

***Abstract*—NASA's Advanced Exploration Systems Autonomous Mission Operations (AMO) project conducted an empirical investigation of the impact of time delay on today's mission operations, and of the effect of processes and mission support tools designed to mitigate time-delay related impacts. Mission operation scenarios were designed for NASA's Deep Space Habitat (DSH), an analog spacecraft habitat, covering a range of activities including nominal objectives, DSH system failures, and crew medical emergencies. The scenarios were simulated at time delay values representative of Lunar (1.2-5 sec), Near Earth Object (NEO) (50 sec) and Mars (300 sec) missions. Each combination of operational scenario and time delay was tested in a Baseline configuration, designed to reflect present-day operations of the International Space Station, and a Mitigation configuration in which a variety of software tools, information displays, and crew-ground communications protocols were employed to assist both crews and Flight Control Team (FCT) members with the long-delay conditions. Preliminary findings indicate: 1) Workload of both crewmembers and FCT members generally increased along with increasing time delay. 2) Advanced procedure execution viewers, caution and warning tools, and communications protocols such as text messaging decreased the workload of both flight controllers and crew, and decreased the difficulty of coordinating activities. 3) Whereas crew workload ratings increased between 50 sec and 300 sec of time delay in the Baseline configuration, workload ratings decreased (or remained flat) in the Mitigation configuration.**

TABLE OF CONTENTS

1. INTRODUCTION 1
2. TEST ENVIRONMENT AND TIMELINE 4
3. EXPERIMENT DESIGN 8
4. MEASURES OF PERFORMANCE 11
5. TASK COMPLETION ANALYSIS 13
6. TIME DELAY AND WORKLOAD 13
7. TEAM COORDINATION.....14
8. COMMUNICATIONS ANALYSIS.....16
9. CONCLUSIONS AND FUTURE WORK.....17
10. ACKNOWLEDGEMENTS.....18
REFERENCES.....19
BIOGRAPHIES.....20

1. INTRODUCTION

For the last 50 years, NASA's crewed missions have been confined to the Earth-Moon system, where speed-of-light communications delays between crew and ground are practically nonexistent. The close proximity of the crew to the Earth has enabled NASA to operate human space missions primarily from the ground. This "ground-centered" mode of operations has had several advantages: by having a large team of the people involved on the ground, the on-board crew could be smaller, the vehicles could be simpler and lighter, and the mission performed for a lower cost.

Table 1. Human spaceflight destinations in the Solar System, approximate distance from Earth, and approximate one-way light time delay.

Destination	Earth Distance (km)	1-way Time delay (s)
Lunar	38,400,000	1.3
NEOs (close)	variable	variable
Mars (close)	545,000,000	181.6
Mars (opposition)	4,013,000,000	1337.6

NASA is now investigating a range of future human spaceflight missions that includes a variety of Martian destinations and a range of Near Earth Object (NEO) targets. These possibilities are summarized in Table 1. The table shows the approximate distance between the

U.S. Government work not protected by U.S. copyright

destination and the Earth, where the control center will be located, and the one-way light-time delay between the destination and Earth.

As is evident from Table 1, future missions will be of much longer duration, and put crews much further from Earth, than today's missions. Accordingly, NASA has recently funded a number of projects to develop and test operations concepts for these future missions. Table 2 summarizes these projects, some of which have included studies of the impact of time delay. We will briefly describe the previous projects here, and refer the reader to the cited studies in the table for more details.

The NASA Extreme Environment Mission Operations (NEEMO) missions [3,8] are conducted at the Aquarius undersea habitat with mixes of astronaut and scientist crews. Extra-vehicular activities (EVA) involve divers, submersibles, and spacecraft analogs; EVA objectives included construction and science tasks. The Desert Research and Technology Studies (DRATS) conduct field tests involving analog spacecraft and habitats; EVA activities focus on science tasks such as gathering and analyzing geological samples [1]. The Houghton Mars Project facility in Nunnavit Territory, Canada focuses on activities ranging from science EVAs to robotic experimentation including drills and mobile robots [8]. Finally, the Mars 500 105 day experiment [8,9] was primarily a living and working experiment with a brief simulated EVA. All experiments included some quiescent activities, but none involved systems failure simulations.

Table 2. Previously conducted investigations of the impact of time delay on spaceflight, compared to the AMO study.

Analog	Year	Time delays	Range of activity ; variations
Haughton Mars Project	2008	N/A	EVA, quiescent; nominal
Mars 500 105 day study	2009	0; 20 minute	EVA, quiescent; nominal
NEEMO 13	2009	0; 20 minute	EVA, quiescent; nominal
NEEMO 14	2010	0; Twice daily	EVA, quiescent; nominal
Desert RATS	2010	0; Twice daily	EVA, quiescent; nominal
NEEMO 15	2011	0; 50 second	EVA
AMO	2012	1.5 sec, 5 sec, 50 sec, 5 minute	Quiescent; nominal, systems failure, medical emergency

While not an operations study, a study of remote medical operations [7] initial assessed whether a communication delay can impact remotely guided collection of ultrasound images. Choosing a communication delay experienced during lunar missions, the investigators demonstrated that increasing the communication delay up to 5 seconds did not impact a remote guidance expert's ability to guide non-ultrasound experts to collect high quality ultrasound images.

On next-generation deep-space missions, crews will have to operate much more autonomously than they do today. A higher degree of crew autonomy represents a fundamental change to mission operations. Enabling this new operations philosophy requires a host of protocol and technology development. To address these issues, NASA's Autonomous Missions Operations (AMO) project charter is to provide operational guidelines and requirements for next-generation crewed missions that will experience significant time delay between mission control and the flight crew. Specifically, AMO addresses the following question: How should mission operations responsibilities be allocated between ground and the spacecraft in the presence of significant light-time delay between the spacecraft and the Earth?

To begin addressing this question, an experiment assessing crew-ground interaction and operational performance was performed in May and June of 2012 in NASA Johnson Space Center's Deep-Space Habitat (DSH) [10,18], an Earth-analog of a workspace and living area that might house a crew during the transport and surface phases of a deep-space crewed mission. Crews consisting of a commander and three flight engineers followed a two-hour mission timeline populated with activities representative of those that might occur during a typical day in the quiescent (cruise) phase of a long-duration space mission. Crews were supported by a small flight control team (FCT) consisting of eight console positions located in the Operations Technology Facility (OTF) in the Christopher Kraft Mission Control Center at Johnson Space Center. The two-hour mission timeline was performed repeatedly under varying conditions:

- A simulated time delay between the ground and the vehicle of low (1.2 or 5 seconds), medium (50 seconds), or long (300 seconds) duration.
- Either no unexpected events (nominal), multiple spacecraft systems failures (off-nominal systems), or a crew medical emergency (off-nominal medical).
- One of two mission operations configurations. In the Baseline configuration, conducted first, the flight control team and crew performed their nominal and off-nominal tasks with support tools, interfaces, and communications protocols similar to those in use for International Space Station operations today. In the Mitigation configuration, crews and FCT members had access to an advanced suite of operations support tools and mission support technologies that we hypothesized would enable the crew to carry out nominal and off-nominal mission operations with greater autonomy and with enhanced crew-ground coordination capability under time delay.

The AMO study complements and extends previous studies of time delay in ground-based analog environments in a variety of ways. The AMO study is the first of the studies in NASA's Earth-analog environments to examine the effects of time delay in an operational environment that:

- Exclusively utilized highly experienced NASA flight controllers and astronauts as study participants.
- Achieved at least a medium level of mission operational fidelity (as rated by the participants).
- Exclusively employed operations products (plans and procedures) like those used in crewed missions today.

In addition, the study was conducted on much shorter timescales (hours) than previous studies (days or weeks), allowed the experiment to incorporate a considerably wider variety of conditions than previous ground-analog studies, and more systematically manipulated the factors of time delay, level of autonomy, and type of scenario. Since the experiment was staffed directly by NASA International Space Station and Space Shuttle flight controllers and astronauts, we made sure to solicit extensive written feedback from participants, both at the end of each run and following all runs, yielding a rich database of observations and expert opinions on the effects of time delay and the impact and usefulness of our mitigation tools. In addition to this written feedback, we collected data on several objective (e.g., task completion time and accuracy), and subjective (e.g., rated workload) measures of performance, along with written explanations of virtually all subjective ratings (e.g., if you rated your workload as "5" on the run just completed, why did you select that rating?). Consequently, we were able to take an integrative approach to data analysis and interpretation; for example, by using participants' written comments to inform our interpretation of empirical patterns in the objective and subjective measures of performance.

The roles and responsibilities of the crews in our study differed fundamentally from those of the FCT. Crewmembers were the primary "doers", responsible for performing most of the procedures associated with their assigned activities, and completing troubleshooting procedures in response to system failures and medical emergencies. While FCT members did play an active role in some of these procedures as well, overall their role was more supportive, advising and guiding crewmembers as they went about their activities. This is due to the fact that all activities in the experiment timeline were 'hands-on', and could not be completed solely with ground commanding. From an information processing perspective, the different responsibilities of crew and FCT suggested that ground personnel might put a high priority on seeking out and processing information sources pertaining to crew activities and progress, in that way maintaining as high a level of situation awareness as possible concerning what the crew was doing, and how well they were doing it. On the other hand, engaging in information acquisition activities to "stay on top of" the activities of FCT members was not as high a priority for the crew. A priori, therefore, crew and ground responses and assessments of the impact of time delay might be expected differ. However, until we ran our study, such differences belonged solely to the realm of conjecture. Our approach to both experiment design and data collection enabled us to systematically compare results and findings between crewmembers and FCT members to identify and quantify such differences, leading to a better understanding of the impact of time delay from the two perspectives.

The second goal of our study was to evaluate the impact of several advanced technologies and decision-aiding tools that we provided in the Mitigation configuration. An important aspect of this evaluation involved comparing objective and subjective measures of performance between the Baseline and Mitigation configurations at the different time delays. Again, there were several a priori reasons to expect the tools would have a positive impact. In Baseline, the only channel for crew-ground communication was the voice loops. Voice communications, of course, must be attended to in real time, and if a communication act is partially unattended or misunderstood, the only way to achieve clarification is to request a repeat of the communication. When a communication act is misunderstood under significant time delay, the round-trip time delay involved in receiving a repeat may well discourage the receiver from making the request at all, meaning that the original communication remains misunderstood. Some such problems could be eliminated in the Mitigation configuration, where additional channels of communication are available (shown in Figure 7). When a communication arrives in written form, e.g. via a texting tool, the receiver can process it "after the fact" with no ambiguity regarding the content. Thus, we would expect the texting feature to be of significant benefit under time delay.

Additional benefits would also be expected from two additional tools. In Baseline, crew procedures were only available in the form of static Portable Document Format (PDF) files, essentially just ported versions of paper files. Navigating through static depictions of procedures is a notoriously workload-intensive activity, partly because the crewmember must keep track of their progress through the procedure strictly from memory. Furthermore, FCT members have no means of tracking crew progress through their procedures except for voice updates. By contrast, in Mitigation, procedures were available in the form of a dynamic procedure display called WebPD (Figure 8). WebPD contained a focus bar that tracked where a crewmember was in a procedure and progressed through the procedure as the crewmember completed steps. In addition, windows were provided that showed which procedures were currently active, and which had been completed. Finally, WebPD was shared over the air-ground link, rendering it

U.S. Government work not protected by U.S. copyright

viewable by all crewmembers and FCT members. This general availability allowed crewmembers to keep track of each others' activities, and enabled the FCT to track the crew's progress through a procedure without resorting to voice or text calls. Another tool, Advanced Caution and Warning (ACAWS), automated two important aspects of Fault Detection, Isolation, and Recovery (FDIR) activities, namely, the initial diagnosis of the source of a failure (depicted in an intuitive fashion on a graphical user interface), and an automatic recommendation of appropriate troubleshooting or recovery procedures. In Baseline, by contrast, both crew and ground had to diagnose the source of failures by integrating information from the legacy caution and warning system (i.e., through failure messages and alerts) and to make their own determination as to which procedure to follow. Without ACAWS, the crew is more dependent on ground expertise to make these decisions, rendering FDIR activities more impacted by time delay.

In summary, the Mitigation configuration provided:

- Tools allowing the crew to visualize spacecraft telemetry and issue commands from procedure displays.
- Tools allowing flight controllers to track procedure execution status across time delay.
- Advanced caution and warning tools to automatically isolate faults and recommend procedures based on vehicle configuration.
- A texting client in addition to voice loops for crew ground communication.

A priori, we hypothesized that:

- Crews would complete less of the timeline as the time delay increased.
- Crew workload would increase as the time delay increased.
- Crew-ground coordination would become more difficult as the time delay increased.
- Crews would complete more of the timeline in the mitigation configuration.
- Crew workload would be lower in the mitigation configuration.
- Crew-ground coordination would be easier in the mitigation configuration.

The rest of the paper is organized as follows. The test environment and activities in the crew timeline are described in Section 2. The experiment design is described in Section 3. The test measurements used to analyze participant workload, coordination, timeline completion, and communications are described in Section 4. The analysis of task completion is described in Section 5. A preliminary analysis of participant workload is described in Section 6. The analysis of coordination difficulty is described in Section 7. The analysis of communications is described in Section 8. Finally, in Section 9, we present our conclusions and discuss future work.

2. Test Environment and Timeline

The Deep Space Habitat

The Deep Space Habitat [10,18] was developed as a functional test article with the flexibility to be used in various configurations. The DSH is comprised of the Atrium/Hab, Hygiene Module, Lab, Deployable Porch Ramp, and Dust Mitigation Module (aka, Airlock). For the purposes of the AMO project, the Lab was the only section of the DSH actively used by the crew, with the Dust Mitigation Module being used for stowage. The Lab module of the DSH is divided into eight pie-piece segments, labeled A-H. The Lab is outfitted with a General Maintenance Work Station (GMWS), a Medical Operations Work Station (MOWS), and a Tele-Robotics Work Station (TRWS) that controlled a camera on the exterior of the Lab. The three workstations - GMWS, MOWS, and TRWS - each have their own segment so as not to interfere with operations occurring at the others. DSH subsystems employed in the AMO experiment are described below.

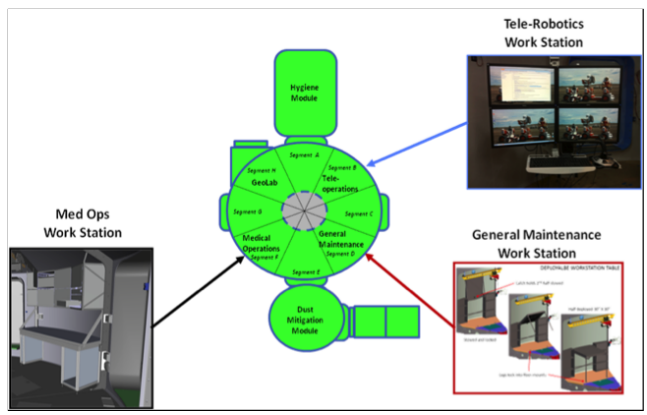


Figure 1. The Deep Space Habitat internal volume layout and key workstations.

DSH Power

The primary power used by the DSH is 120 Vac supplied from a variety of sources. Secondary power sources of 120 Vac, 28 Vdc, and 120 Vdc were also available for use. The power system schematic is shown in Figure 2.

Sensors

Instrumentation system sensors provided data to each of the DSH modules and airlock subsystems to provide insight into system performance. These sensors were powered by the DSH power.

External Camera

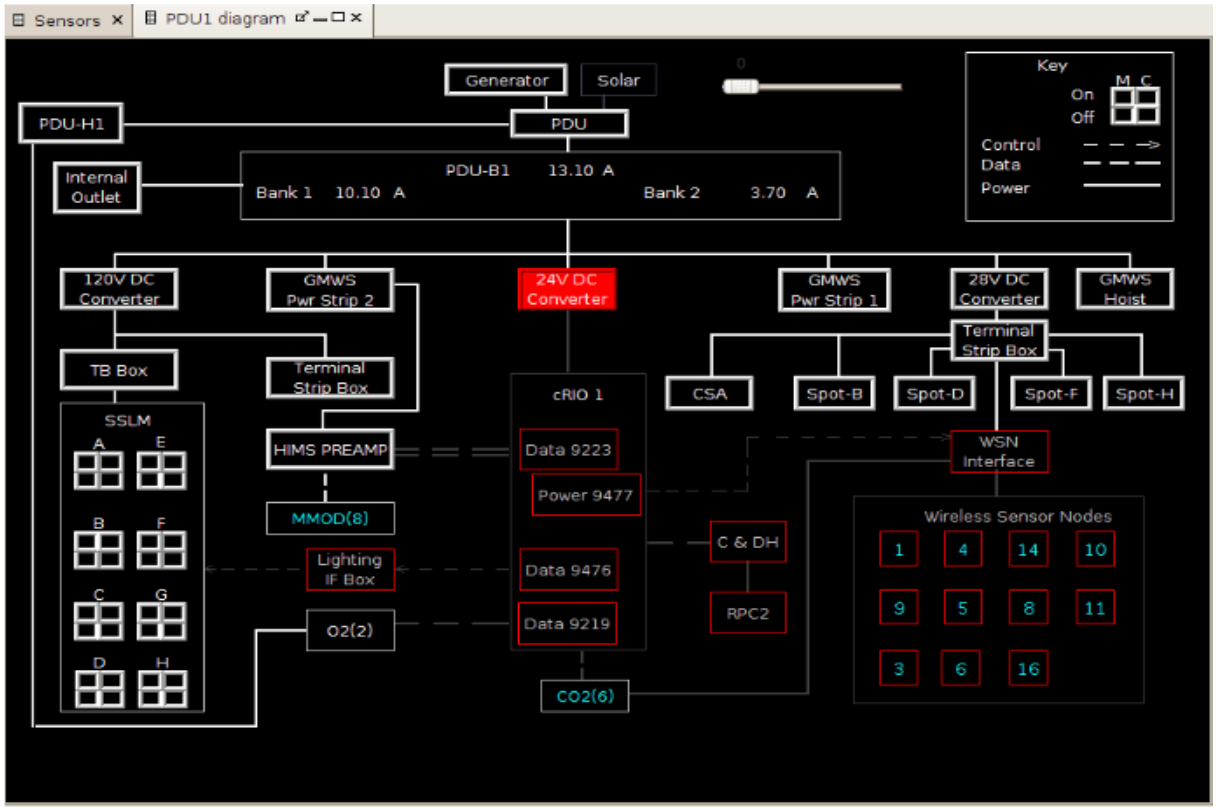


Figure 2. DSH Electrical Power System schematic, also showing ACAWS User Interface.

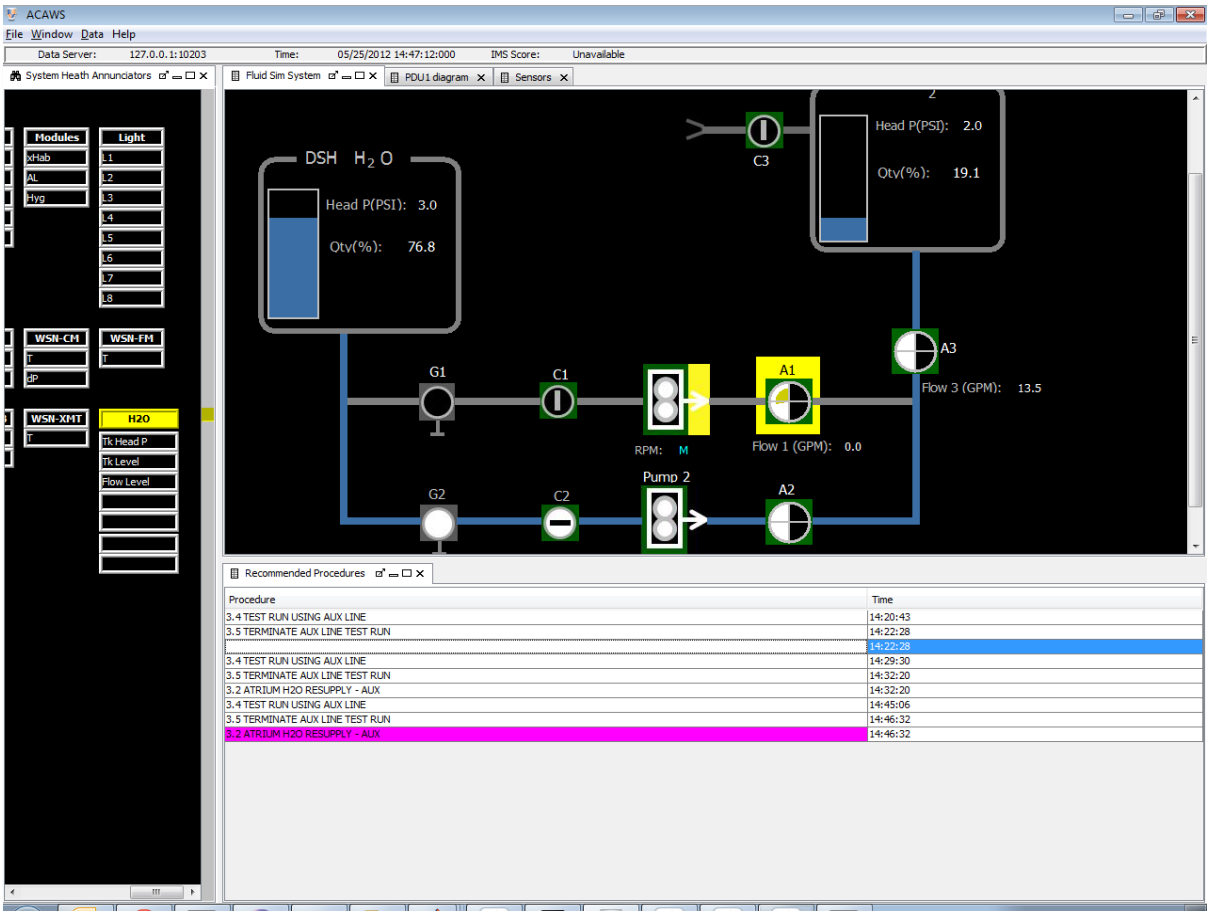


Figure 3. DSH fluid transfer schematic, also showing fluid transfer User Interface.

The DSH is equipped with a Sony SNC RZ30N network equipped pan-tilt-zoom camera. This camera is mounted between segments G and H on the outside of the DSH. The camera was integrated with the DSH avionics in order to be commandable from inside the DSH. The camera was also integrated with the DSH avionics so that all acquired camera images were placed on the file system of the TRWS, and downloadable from there to computers in the FCT.

Fluid Transfer System

For the AMO simulations, a water transfer activity was simulated on a laptop in the Lab to simulate the transfer of water between a virtual DSH primary water supply tank and the Atrium Water tank used to water the onboard plants. Anin valves ("A" valves below) had multiple possible positions. "C" valves were computer controlled and could be commanded by the FCT or the crew through the laptop. "G" valves were Gate Valves that had to be commanded through a hardware switch simulated on the GeoLab workstation computer and could not be commanded from the ground. The design of this system is intentionally more complex than required if it were a real system to make failure scenarios on par with DSH electrical failures for experimental design purposes. The fluid transfer system is shown in Figure 3.

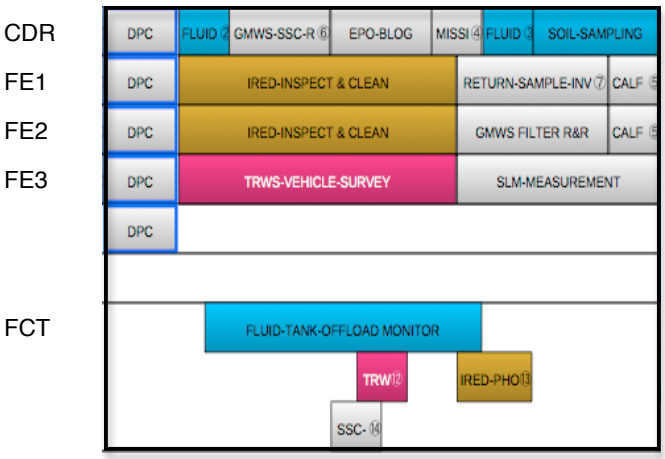


Figure 4. The 2 hour timeline of all tasks scheduled for execution by the crew for the AMO experiment. The first four rows indicate activity performed by each crew person, the last bands show required coordination by the FCT for each activity.

Timeline

The experiment employed variations of a timeline of activities that the crew needed to complete. For the simulation "initial conditions", the vehicle was returning form an asteroid and was in a "quiescent" operational mode, meaning there are no significant, complex or dynamic operations scheduled (i.e. no burns or other maneuvers were planned for the day). The vehicle was in a nominal configuration except for some designated conditions listed below, and there were no previous major systems failures.

The crew's timeline consisted of 12 activities of varying duration during a two-hour period, and is shown in Figure 4. In the Baseline configuration, these activities were preceded by a 10 minute schedule-prepwork activity and a 15 minute Daily Planning Conference (DPC) activity, in which the flight control team briefed the crew on the specifics of the day's timeline. In the Mitigation configuration, these activities were merged into a single schedule-prepwork activity of 25 minutes. The most important information passed up during the DPC were parameters for the atrium tank fluid fill (system set up conditions, target fill level and estimated fill duration).

Atrium Tank Fluid Fill: The atrium fill condition for the experiment was as follows: "The plants in the Atrium which require the most amount of water are starting to show health degradation and thus the crew has been asked not to take any fresh vegetables from the plants. The working theory is that the Atrium plant irrigation cycles have been using higher amounts of water than expected which dropped the water level in the Atrium tank. This caused the concentration of automatically added additives (fertilizer, macronutrients, etc) to be increased in the irrigation water which in turn has affected the plants. The Atrium tank will be resupplied with fresh water today during the fluid fill activity to reduce additive concentration within the tank. Once the tank is filled, the water should sit for 2 hours to allow the additives to fully mix with the new water. After those 2 hours, the plants will be watered for 8 hours. Plant watering must be complete prior to pre-sleep activities." These time constraints were selected so that there was some slack in the schedule (~ 1 hour) to water the plants before sleep but not much slack time, making this a high priority activity.

Vehicle Survey: The vehicle survey condition for the experiment is as follows: "The crew reported late in their day yesterday hearing unusual noises on one side of the DSH. No onboard sensors have indicated any off nominal vehicle system issues. An external vehicle survey has been scheduled today to view the external area of the DSH where the crew thinks a possible meteor strike may have occurred. This survey will be conducted by the crew, with ground assistance, using a robotic camera system mounted outside the spacecraft." This was the highest priority activity in the timeline, and per Flight Rules, the survey must be conducted as soon as possible within 24 hours of a suspected impact.

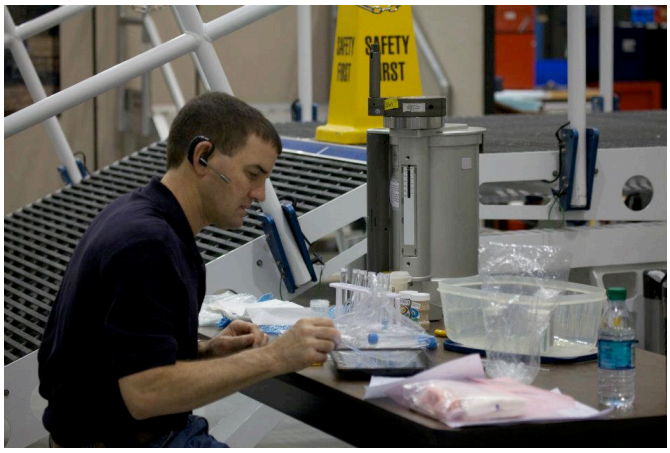


Figure 5. Soil pH sample task.

Soil pH determination: The activity description read: "Prior to the plants being watered, soil pH should be measured for the plants in question to get a baseline reading of the plant's growing condition. If any soil pH is found to be outside of the acceptable range, the test for that plant will be repeated again tomorrow and it should not be harvested for food to prevent further stress on the plant. The areas of high additive concentration will need some time to be broken down. It is expected that it may be a minimum of 48 hours before the full range of fresh fruits and vegetables will once again be available on the menu. Some sections may be available sooner, but that will depend on what the data analysis shows." This activity is shown in Figure 5.

Interim Resistive Exercise Device (iRED) Inspection and Cleaning: The iRED activity description was as follows: "At the end of the last crew day, the crew reported some grinding coming from the IRED canister. The crew will disassemble, inspect and clean the canister at the start of the sim timeline. Inspection photos will be downlinked to the ground for analysis which takes 20 minutes. The crew must wait for FCT "Go" before performing any exercise. Due to previous failures, this is the only piece of resistive exercise equipment available onboard." It is unrealistic for any ground analysis to take only 20 minutes; the duration was shortened for the scenario to fit within the 2-hour simulation schedule. The point is for the crew to send data down to the ground in time for ground to give the crew 'go' to exercise. This activity is shown in Figure 6.

Return Sample Inventory: This activity required inventory and sorting of asteroid samples being returned to Earth. The condition read: "There is a concern that the samples taken on the fifth day of operations at the asteroid were contaminated. Payloads has requested the crew examine those samples again and send some additional data for comparison against the initial assessment of those samples."

Space Station Computer (SSC) Hard Drive Troubleshooting: The activity description read: "The last time the crew attempted to use a specific SSC it could not access the hard drive. The CDR has an activity today to attempt to troubleshoot."

Figure 6. iRED inspection and cleaning task.

Missing Item Search: The activity description read: "A few days ago an Ovoid Canister required for an Environment Control and Life Support Systems (ECLSS) onboard activity could not be found in the location documented in the onboard inventory system. The ground would like the crew to spend a few minutes looking for the lost item. If the item is found the crew reports the location and FCT provides a new storage location."

Air Filter R&R: This activity required replacement of four DSH Air Filters. Per Flight Rules the air filters should be replaced every 50 days, but are certified for 75 days of operation.

Bicep and Calf measurement: Measuring the calf and bicep muscle for atrophy. This activity was designed to be representative of a nominal medical procedure

Education and Public Outreach (EPO) Blog: Crew members composed a blog entry about their day aboard the spacecraft and the impact of communication time delay. These blogs were provided to JSC Public Affairs Office (PAO) and posted on a public NASA Website.

Sound Level Measurement: This activity required measurement of ambient sound levels within the DSH. Per Flight Rules, sound level meter readings are required every 150 days. It has been 145 days since the last time this activity was complete and it is currently scheduled for today.

PAO Event: A time critical event, which served as a milestone to reach by the end of the two-hour simulation period.

Three 'get-ahead' tasks were also provided in the event the crew had extra time or an activity needed to be abandoned and replaced; an equipment inventory task, Just-in-time training videos, and additional bicep and calf measurements. Activities required one or two crew; some required support by the flight control team. For example, the iRED, DSH Backside Inspection, Sound Level Meter, Plant Soil pH, Calf length measurement, and sample inventory activities all required data collected onboard the DSH be transferred to the FCT. The iRED, Plant Soil Ph, and Fluid transfer activities all required the FCT to coordinate with the DSH crew during at least part of the activity. This coordination ensured that the crew and the FCT would communicate periodically, even in a nominal scenario.

3. EXPERIMENT DESIGN

Flight Control and Crew Roles

Each AMO crew was comprised of four members, corresponding to the current NASA reference mission crew sizes. Table 3 describes the four positions and the responsibilities allocated to them for the AMO experiment. A total of four distinct crews participated in the experiment to ensure that learning did not skew the quantitative results. For each crew, the Commander (CDR) was a previously-flown astronaut; the remaining three crewmembers were experienced trainers and flight controllers from MOD (Mission Operations Directorate).

Table 3. Crew Position Descriptions.

Commander (CDR)	Oversees mission operations, coordinates priorities with FCT.
Flight Engineer (FE-1)	Backup to commander; performs day to day vehicle maintenance and sample inventory
Flight Engineer (FE-2)	Performs day to day vehicle maintenance
Flight Engineer (FE-3)	Primary robotics operator and Chief Medical Officer

The AMO Flight Control Team (FCT) consisted of eight console positions. Table 4 shows the console positions and the technical topics assigned to the console. The "CAPCOM" and "FLIGHT" console names are legacy titles. The rest of the consoles were named after well-known asteroids. The CAPCOM console was staffed by non-astronaut certified International Space Station (ISS) CAPCOMs with the exception of two runs that were staffed by an astronaut certified ISS and Shuttle CAPCOM, and one run that was staffed by a an experienced Shuttle flight controller.

Table 4. Flight Control Team Positions.

Console Name	Current Discipline Corollary
CAPCOM	Air to Ground (A/G) Communication with the crew
CERES	Payloads and science
FLIGHT	Coordination of flight control team, mission success
IRIS	Robotics
JUNO	Electrical and environmental systems
KALI	Planning and Public Affairs
PSYCHE	Biomedical Engineer (BME); Crew medical monitoring
VESTA	General maintenance and computers

The Flight Director console was staffed for five runs in both the Baseline and Mitigation simulations by two certified Flight Directors, one certified for both Shuttle and ISS and one certified for ISS only. The rest of the runs were staffed by experienced role-play Flight Directors who are senior flight controllers from ISS and Shuttle.

The remaining six FCT console positions were staffed with experienced flight controllers from the ISS and/or Station programs. The majority of the controllers had Flight Control Room (or "front room") experience; a few had certifications at the "back room" or support / analyst level. The PSYCHE console was staffed by certified ISS Biomedical Engineers (BMEs) and the KALI console was staffed by certified ISS/Shuttle Planners. The other consoles were staffed with flight controllers from a variety of discipline backgrounds. Since the AMO activities and DSH systems were relatively simple, there was not a need to match a flight controller's technical background with the corresponding AMO console position. The AMO experiment did not employ distinct flight control teams due to limitations on available personnel.

Experiment Parameters

The experiment varied three parameters: the inserted time delay, scenarios involving deviations from the original timeline, and the operations configuration.

The AMO experiments originally used three different time delay values. For technical reasons, results from the low time delay values were confounded with amount of operator training, so the later sections of the paper will focus the presentation on results for runs with one-way light-time delay values of 50 seconds and 300 seconds.

Experiment runs conducted with no planned deviations from the original timeline are referred to as *Nominal*. In addition, experiment runs conducted with inserted failures of both the fluid transfer and electrical systems are referred to as *Systems*. Finally, runs conducted with a crew medical emergency are labeled *Medical*.

Medical Scenarios

Thirty-minutes into the start of a run with a scripted medical failure, a NASA Exploration Medical Capability (ExMC) moderator supporting AMO identified a crewmember to act as the ill astronaut, and subsequently had that person act out the symptoms of the medical condition to the Crew Medical Officer (CMO). During the initial sequence of the scenario, the moderator provided the ill crewmember with the information needed for any questions asked by the CMO (e.g. answers to examination questions and vital sign data). At a certain point, the moderator took over the role of the ill astronaut with that crewmember no longer participating in the scenario. This was done to keep the scenario relatively consistent between crews and, thus, help with comparisons between sessions.

During the Baseline configuration runs the medical scenario was ultimately diagnosed as urinary retention; for the mitigation runs, the problem was a kidney stone. These two failures had similar initial symptoms and resulted in a similar ultrasound diagnosis process.

Systems Scenarios

For simulation cases with DSH system failures, two different equipment failures were injected at 30 minutes, and subsequently, 1 hour 15 minutes, into the simulation run. The first failure introduced was the A3 valve failing to a maximum open value (100% open). The fluid transfer plumbing, tanks, and associated valves, pumps, pressures and tank quantities were entirely simulated in a MatLab model. Interfaces within the model allowed simulation supervisors to inject failures to valves and pumps, overriding any crew or flight controller commanding, resulting in changes to the flow characteristics. The fluid system failure was timed to occur shortly after fluid transfer was initiated, with the Atrium Tank level below 45% full. By 30 minutes into the simulation, the crew had initiated the flow, and typically had moved on to other planned activities when the failure was injected.

The second failure, injected at 1 hour 15 minutes into *Systems* runs, was the 28 Volt Power Converter failing off. In preparation to support such a failure, the 28V converter switch was relocated from Power Distribution Unit PDU_B1 Bank 2 Port 2 to PDU_B1 Bank 2 Port 6. This relocation allowed the simulation supervisor to turn off the power supply to the 28V converter with no indication from the switch position indicator on the DSH Crew Display. This allowed the 28V converter to be turned off by simulation supervisors, but inspection of the DSH Crew Display would still show the Port 2 switch as ON and downstream loads from the 28V converter as offline. This was also effective in causing the software to recognize the condition as a fault and trigger the appropriate failure messages. The timing of this failure was again consistent for all systems cases, and was timed to occur with sufficient time left in the simulation to complete FDIR procedures. The timing also generally lined up with completion of the Fluid Transfer operation.

Both systems failures were designed in such a way that the root cause of the failure was not completely obvious when the failure occurred, and required one or more diagnostic steps in order to positively identify the source of the problem and determine the appropriate isolation and recovery procedures.

Baseline and Mitigation Operations configurations

The *Baseline* operations configuration was designed to be similar to the manner in which the International Space Station is operated today. The crew has the primary responsibility for conducting each activity, supported by the ability to monitor data from all spacecraft systems on a suite of crew displays. However, the crew had limited in-depth knowledge of DSH systems and their operation. By contrast, the FCT included specialists with in-depth knowledge of each DSH system and its operating characteristics; the crew could turn to this expertise to support them, particularly during the off-nominal Systems runs. However, at the longer time delays, the FCT could only respond to crew questions after a significant time delay. Similarly, FCT could monitor spacecraft telemetry, but not in real time. Recall that all crew and flight control communications could take place only via voice loops.

Each spacecraft subsystem in the DSH came with a legacy Caution and Warning system that provided only limited machine-based fault management assistance to crew and FCT alike. The system reported faults only if single test parameters (sensor outputs) were determined to be outside pre-specified tolerances (limits). This allowed the crew or the flight control team to determine that a fault has occurred, but provided no more assistance with the additional steps that are typically required to diagnose, isolate, and recover from systems malfunctions.

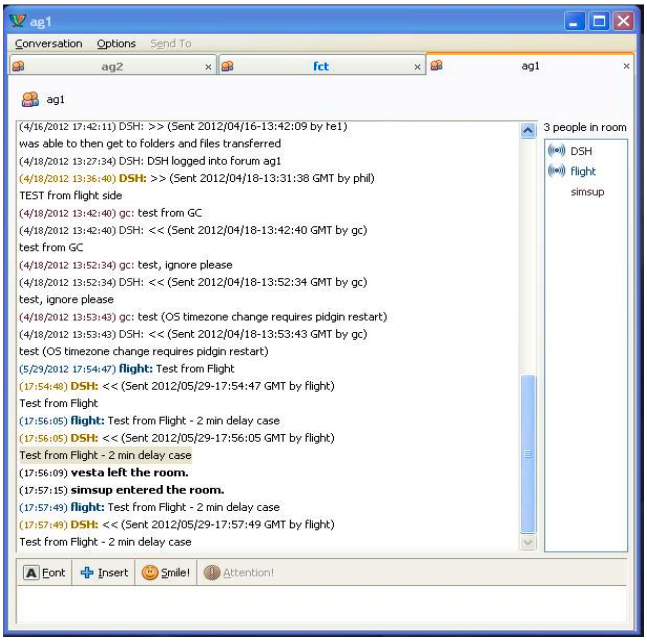


Figure 7. Texting client used by crew and flight controllers.

Finally, the crew's plans, system-specific procedures, and other spacecraft specific knowledge, exist in the form of office documents. Thus, if the crew has questions about the significance of an activity that may need to be skipped, or if the flight control team wishes to know what step of a procedure the crew is on, this coordination must take place over voice loops.

The *Mitigation* operations configuration differs from the Baseline configuration in several key respects. First, a texting[1] client provided an additional communication channel between the flight control team and the crew. The flight controllers and crew had two air-ground chat rooms, and a third chat room was reserved for the flight control team. The crew initiated all air-ground text messages, and texting was intended for only non-urgent or non-emergency messages. The text interface is shown in Figure 7.

Second, Advanced Caution and Warning (ACAWS) software technologies provided both automated detection and isolation of many faults, and automated recommendations of fault isolation or recovery procedures. The AMO experiment employed the Testability and Engineering Maintenance System (TEAMS) [14], a

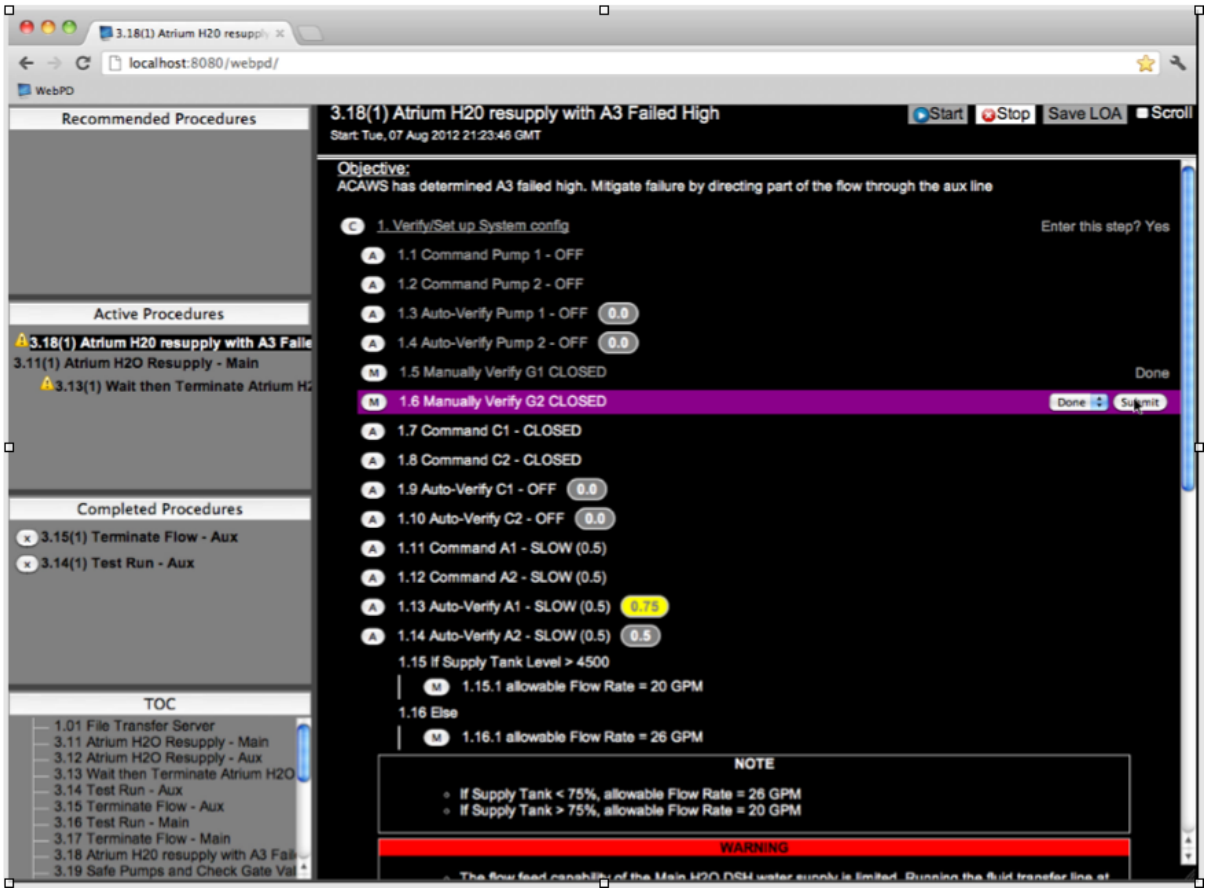


Figure 8. WebPD, the electronic procedure display and procedure execution tracking tool.

[1] The aviation community uses a texting tool called Datalink.

Commercial Off The Shelf (COTS) tool, which was applied to detect faults in the Electrical Power System (EPS). TEAMS is a model-based system; the model captures a system's structure, interconnections, tests, procedures, and failures, which is the relationship between various system failure modes and system instrumentation. More precisely, a pass-fail test (performed on the data from instruments) provides evidence of the possible failure of one or more systems 'upstream' of the test. TEAMS determines the root cause (failed components and their failure modes, the "bad" components in the TEAMS vernacular) using multiple test results. When the test results cannot uniquely identify a single failed component, TEAMS provides a list of possibly failed components (the "suspect" set). Customized schematic displays of the EPS system rendered the good, bad and suspect output of TEAMS for use by the flight controllers and crew during the AMO experiment; the UI is shown in Figure 2. TEAMS was part of the Ares 1-X launch vehicle Ground Diagnosis Prototype [17] and the TacSAT3 satellite Vehicle System Management (TVSM) experiment [13].

Third, the procedures for operating spacecraft systems and performing tasks were presented using an electronic interface called WebPD, shown in Figure 8. WebPD incorporated a focus bar, allowing the crew to track their place in a procedure. The crew could issue commands to spacecraft systems from WebPD. Procedure steps often required reading system data values or checking limits; WebPD 'listens' to all system data, and these are incorporated in the WebPD interface. ACAWS could send messages to the WebPD, prompting the crew to perform a procedure. WebPD could be configured to automatically issue instructions, or act as an automatic scripting engine.

WebPD procedures are stored in Procedure Representation Language (PRL), a derivative of XML [11], and developed in a graphical environment called the Procedure Integrated Development Environment [6]. PRL has been developed over many years by NASA. PRL and a predecessor of WebPD have been used in previous simulations of mission operations environments [5, 12].

Finally, WebPD status was shared over the air-ground link, so that the flight control team could see what procedures were executing, and what procedure step the crewperson running a procedure was presently executing.

It is apparent that each of the elements of the mitigation configuration are complementary. However, only in the cases of ACAWS and the WebPD was there tight integration between technologies in the mitigation (i.e. ACAWS notification of procedures to run, which are then shown to the crew in WebPD) .

The specific Mitigation configuration components are summarized in Table 5.

Table 5. Summary of Baseline and Mitigation Configurations

	Baseline	Mitigation
Communications	Voice	Voice + Texting
Fault Management	Limit Checking	Advanced Caution and Warning
Procedures	PDF	Electronic Procedures
Situational Awareness	Voice	Voice + Shared Electronic Procedures

Descriptions of Runs

For each time delay (50 and 300 seconds), five scenarios were conducted: one *Nominal*, two *Systems*, and two *Medical*. This group of scenarios was repeated for both the Baseline and Mitigation configurations.

Crews were assigned scenarios in such a way that:

- Each crew experienced at least one Nominal, one Systems, and one Medical scenario.
- Each crew experienced both 50 and 300 second time delays.
- Each crew experienced the same combination of time delay and scenario in both the Baseline and Mitigation configurations.

4. MEASURES OF PERFORMANCE

Several types of data were collected for analysis. Since the main purpose of the study was to assess how flight controllers and crew were impacted by the time delays and configuration (Baseline or Mitigation), the AMO team created surveys consisting of subjective ratings and comments to evaluate performance. Data on the number of activities completed and procedure execution logs were collected to provide quantitative analysis. Voice and texting communications were logged in order to provide quantitative analysis of the ways that the team coordinated under the various test conditions. These data collection methods are described further below.

Objective Measurements

Two objective measurements provided insight into the impact of time delay, and differences in configuration. The first of these measurements is task completion. As tasks were started and completed, the crew would notify the KALI flight controller, who in turn recorded this information. As a result, start time, end time, and activity completion for every experiment were recorded for later analysis.

The second of these measurements is communications between the flight controllers and crew. In both the Baseline and Mitigation configurations, each voice call start time and end time was recorded. In the Mitigation configuration,

each text message was logged and archived. This permitted quantitative analysis of how team communication varied.

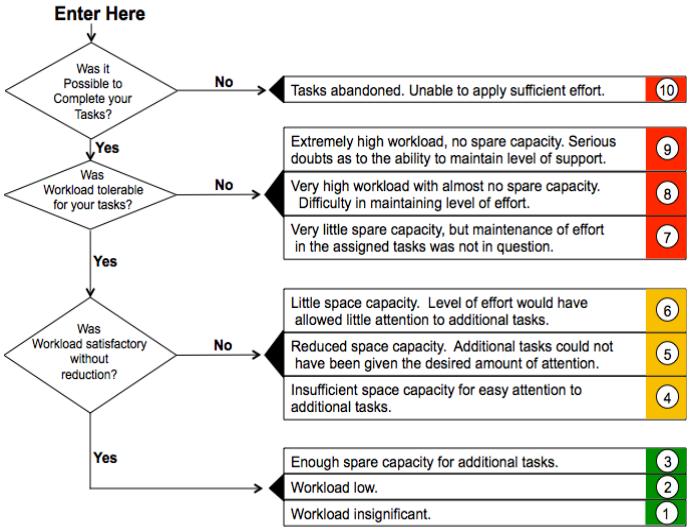


Figure 9. The Bedford workload rating questionnaire.

Subjective Measurements

Immediately following each run, each participant completed an electronic questionnaire. The first order of business on the questionnaire was to select a workload rating for the just-completed run on a slightly modified version of the Bedford workload rating scale. Shown in Figure 9, the Bedford asks participants to rate their workload on a scale from 1 to 10, with values from 1 to 3 associated with the lowest "green" category (workload satisfactory without reduction), values from 4 to 6 associated with the intermediate "yellow" category (workload unsatisfactory without reduction), and values from 7-9 associated with the highest "red" category (workload intolerable for your tasks). Bedford was chosen because Bedford is an "anchored" scale; each point on the scale is associated with a clearly specified selection criterion, based on assessments by operators of how much spare attentional capacity they thought they would have had to perform additional tasks, should any have been imposed. This stands in contrast to unanchored scales, like NASA's TLX, that leave the criteria for selecting one value over another much more arbitrary [15,16]. From an operational evaluation perspective, the demarcation of workload into a three-colored color scheme allows analysts to determine whether an overall operational environment produced a satisfactory (Green zone) versus unsatisfactory (Yellow/Red) level of workload for its operators, and is therefore in need of additional tool development, tool improvement, or other alteration to further reduce workload, or the current ops environment yielded a satisfactory workload level (green area), rendering further modifications to the ops environment unnecessary.

The workload rating was followed by 10 questions that were each answered by selecting one value from a five-point rating scale. Several questions targeted crew-ground coordination issues (e.g., "In the run you just completed, how difficult was it to coordinate activities with crew/ground" (1 = very easy to coordinate, 3 = moderately difficult to coordinate, 5 = very difficult to coordinate, 6 = Not Applicable)". Other questions asked for an explicit rating of the impact of time delay on a specific operation (e.g., "assuming the run you just completed included a systems malfunction, please rate the impact of the time delay on your ability to work the malfunction (1 = no impact, 3 = moderate impact, 5 = strong impact, 6 = run contained no systems malfunction or I was not involved in working the malfunction).

In addition to these subjective metrics, comments from the observers were solicited after each rating to provide better understanding of their ratings choice, and to acquire further insight into their view of what happened on the run and why. For example, Question 7.1 was worded as follows: "Assuming the run you just completed contained a systems malfunction, please rate the impact of the time delay on your ability to work the malfunction (1 = no impact, 3 = moderate impact, 5 = strong impact; 6 = the run contained no systems malfunction or I was not involved in working the malfunction). The following question then asked for written comments to clarify respondents' choice: "If you responded to question 7.1 with a numerical rating (i.e., the run contained a systems malfunction and you had some involvement in it), please explain your choice. If you rated the impact of time delay on malfunction handling as minor, (1 or 2), was that because the time delay was small, or the software tools (i.e., ACAWS) and communications protocols provided effective mitigation, or coordination with crew wasn't necessary or important? If you rated the impact as moderate or strong, (3 or more), how did the impact manifest itself? In greater difficulty coordinating activities with crew? In disruptions of voice loops with crew? In maintaining a shared "mental model" of the situation with crew? In all (or none) of the above?"

At the completion of each participants' final run, after completing the "after-run" survey, they proceeded to complete a second "wrapup" questionnaire. The "wrapup" survey included a series of questions designed to elicit usability opinions and evaluations of the software tools provided during Baseline (the PDF-based procedure displays and limit-based C&W tools). For each tool, opinions were solicited by asking participants to note three things that they liked about the tool and three things that they disliked about the tool, followed by a more open-ended opportunity to make any additional comments and recommendations for feature improvements. These questions were repeated in the "wrapup" questionnaire administered at the completion of all Mitigation runs, but with evaluations of the tools available during Mitigation (e.g., WebPD, ACAWS, and texting).

5. Task Completion Analysis

On the assumption that our tasks entailed a reasonable level crew-ground interaction, one of our most straightforward hypotheses was that time delay would lower the proportion of timeline activities crews were able to complete.

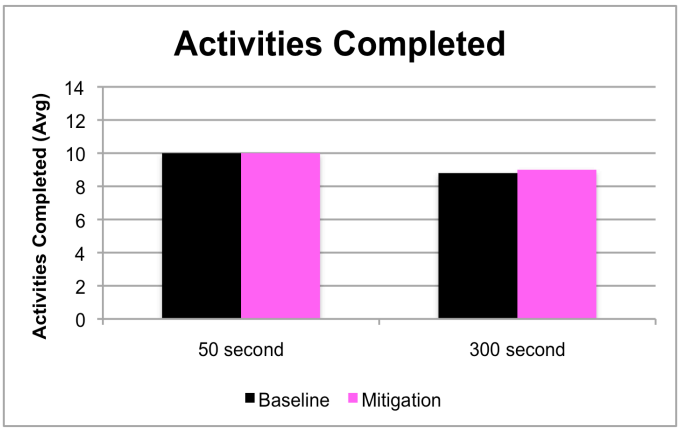


Figure 10. Timeline task completion, averaged across the 5 runs at 50 and 300 seconds.

Figure 10 shows the average number of activities completed over all individual runs at 50 seconds and 300 seconds of time delay. Recall that there are 5 scenarios at each time delay; one Nominal, two Systems, and two Medical, and that the same crew, time delay, and the combinations of crew and scenario selected for Baseline were repeated in Mitigation. There did appear to be a reduction in the number of activities completed as the time delay increased from 50 seconds to 300 seconds; on average, one fewer task is completed at the higher time delay. This reduction is quite modest, however, and there appeared to be no difference in activity completion rates as a result of configuration. Furthermore, recall that the Baseline configuration timeline included two activities (schedule-prepwork and Daily Planning Conference) as compared to a single activity (schedule-prepwork) at the beginning of the timeline. Since these activities were always completed, regardless of run, it may be fairer to say that more activities were completed in the Mitigation configuration than the Baseline configuration. Since the unified schedule-prepwork activity consisted of a single 25 minute block of time, the extra time could have been used to complete the task.

Despite these nuances, the tentative conclusion is that, rather surprisingly, activity completion rates were not impacted to any meaningful extent by either time delay or configuration.

6. Time Delay and Workload

Crew Workload.

Figure 11 shows the average workload ratings of crewmembers as a function of time delay and operational configuration (error bars indicate the standard deviation of the distribution of ratings scores obtained in each condition). Recall that a Bedford workload rating of three or below falls within the "green zone" (workload satisfactory without reduction), whereas ratings of four to six fall in the "yellow zone" (workload unsatisfactory without reduction). As shown by the error bars, most ratings fell in a range between two and six, with three of the four averages straddling the border between "Green" and "Yellow".

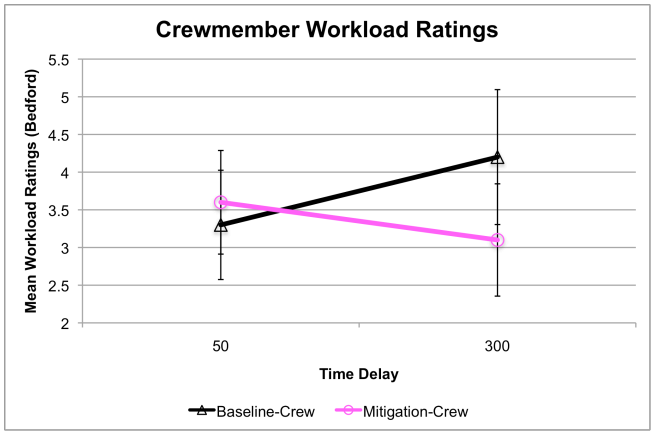


Figure 11. Crew workload, averaged across the 5 runs at 50 and 300 seconds.

The figure also reveals that in Baseline, the average rating fell just above the "Satisfactory without Reduction" (3.25) range at the 50 second time delay and increased to the "Unsatisfactory without Reduction" range (4.1) at 300 seconds. In Mitigation, on the other hand, the average rating *decreased* between 50 and 300 seconds, almost reaching the desirable "Green" zone at 300. This is an interesting pattern that we had not expected. A three-way Analysis of Variance (ANOVA) with crew, time delay (50 versus 300 sec) and operational configuration (Baseline versus Mitigation) as factors revealed no main effect or interaction involving crew, no main effect of configuration or time delay, but a significant interaction between configuration and time delay, $F(1,12) = 10.36$, $p < 0.01$. Individual comparisons revealed that the difference in workload between 50 and 300 seconds was significant in the Baseline configuration ($p < .05$) but not the Mitigation configuration.

FCT Workload.

Overall, workload ratings were considerably lower among FCT members than crewmembers. This is probably because, as we noted earlier, the FCT had more of a supporting role in flight operations than the crews. Indeed, a rank ordering of average workload ratings in the Baseline configuration by FCT position revealed that the ratings for fully four of the eight flight controllers console positions fell in the low (Green) zone. This is largely due to lack of involvement of these operators in most tasks on the timeline. In an attempt to eliminate these floor effects and increase the sensitivity of statistical testing, only the data for the four highest workload positions (FLIGHT, CAPCOM, KALI and CERES) were subjected to statistical analyses. The average workload ratings of just these highest-workload console positions are plotted in Figure 12.

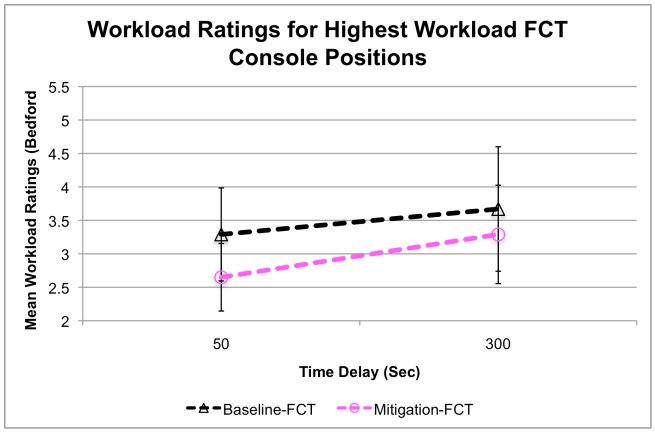


Figure 12. Highest four flight controllers' workload averaged across the 5 runs at 50 and 300 seconds.

In a clear departure from the pattern exhibited by crewmembers, Figure 12 reveals that flight controller workload ratings were consistently higher in Baseline than in the Mitigation, and higher under 300 seconds of time delay than under 50 seconds of delay for both configurations. An ANOVA with crew, time delay, and configuration as factors revealed marginally significant effects of both configuration [$F(1,12) = 3.99$, $p < .07$], and time delay [$F(1, 12) = 4.12$, $p < .07$], and no hint of an interaction.

7. Team Coordination

What factors contributed to the increase in workload for both crew and FCT between 50 and 300 sec in the Baseline Condition, and why did crew workload either stay flat or *decrease* slightly across time delay in the Mitigation condition, but *increase* for FCT members? Determination of workload ratings was completed "open ended"; to avoid any bias toward our experimental manipulations, participants were not supplied with any guidance concerning what features of the operational environment they should consider when determining their ratings. Thus, it was interesting to note that in their explanation for why they selected the rating they did in the Baseline configuration, several crewmembers identified ground coordination issues as a contributing factor; similarly, coordination issues with crew (in the case of FCT members) was a common theme in the comments of FCT members. For example, one crew member noted: *"Time delay made it difficult to do voice comm and still keep your place in procedures since the time is long enough the crew moves onto other tasks while waiting for the [Mission Control Center] MCC to get back in touch for further direction."*

This quote was one of several that pointed out that time delay increased the multitasking demand on crews, as they started new tasks before existing tasks were completed while awaiting feedback from ground on tasks not yet completed. Multi-tasking imposes a number of demands on memory and activity coordination that might be expected to increase workload.

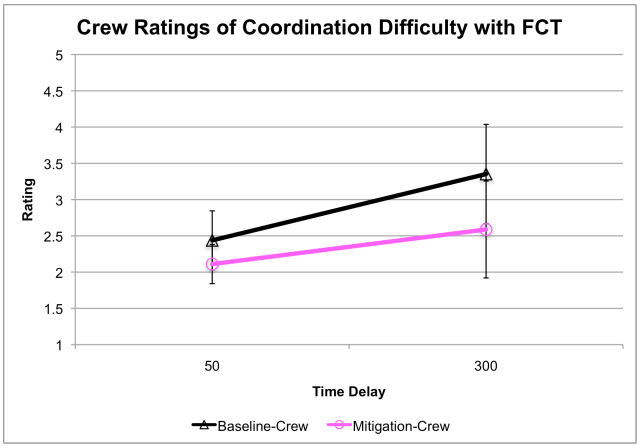


Figure 13. Crew rating of coordination difficulty, averaged across the 5 runs at 50 and 300 seconds.

Another crewmember noted: "*No satisfying feedback that any transmission if [sic] info (voice, files, crew notes) was being received or buffered at the ground in a timely enough manner that it didn't exceed the length of my short term memory. So I had to write info down in case I got a "say again" or "file not received" message back from MCC minutes after I'd dumped the details from my buffer.*"

The second quote indicates that the long time delay condition forced the crew to incorporate additional coordination-related activities that weren't necessary when time delay was short – another obvious candidate for increasing workload. In general, then, the evidence from these (and many other) comments suggests that crew/ground coordination issues were a significant contributor to the increase in workload experienced by crew (and possibly ground) in Baseline when time delay increased from 50 seconds to 300 seconds.

If this hypothesis is true, coordination difficulty itself would be expected to be increase along with time delay. As part of the questionnaire that immediately following each run, both crew and flight controllers were asked the following question:

"In the run you just completed, how difficult was it to coordinate activities with the ground"? (1 = not at all difficult to coordinate, 3 = moderately difficult to coordinate, 5 = quite difficult to coordinate)."

If coordination difficulty contributed to higher workload at the longer time delays, we would expect to see coordination rated as more difficult in the 300 sec condition compared to the 50 sec condition. Figure 13 shows the average and standard deviation of the crew's ranking of the difficulty of coordination with the flight control team; Figure 14 shows the equivalent rankings for the FCT. As before, the flight

controller Figure includes ratings from of the high workload flight control positions only.

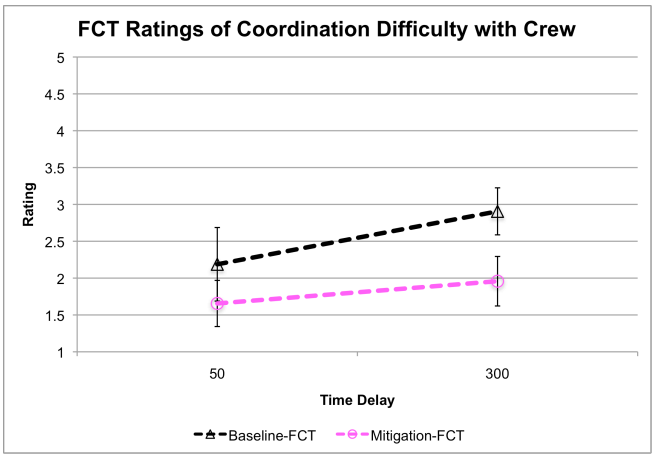


Figure 14. Highest four flight controllers' coordination difficulty, averaged across the 5 runs at 50 and 300 seconds.

The Figures show that ratings of coordination difficulty did indeed increase with time delay, both for the crews and flight controllers. The Figures also reveal that coordination was ranked as easier in the Mitigation configuration than in the Baseline configuration. These observations were supported by statistical analyses. In an ANOVA on crew ratings including crew, configuration, and time delay as factors, the main effects of configuration [$F(1,12) = 9.55$, $p < .01$] and time delay [$F(1,12)$, 7.57, $p<.01$] were significant. For FCT members, the corresponding tests for configuration [$F(1,12) = 14.5$, $p<.01$] and time delay [$F(1,12) = 5.31$, $p<.05$] were also significant.

Still further evidence of a link between workload ratings and coordination difficulty can be found in the fact that in Baseline, the product-moment correlations between workload ratings and ratings of coordination difficulty were considerable, .43 for the crew and .51 for ground. In Mitigation, the correlation was unchanged for FCT at .51 but dropped to .29 for Crew. This pattern would be expected if crews operated more autonomously in Mitigation, rendering coordination difficulty less influential.

While coordination difficulties are implicated in the increase in workload experienced between 50 and 300 sec of time delay, it is also clear from the data that workload was reduced in the Mitigation configuration for flight controllers and for crew, most notably for the crew at the 300 sec time delay. In the following sections, we delve into participants' comments about the usefulness, impact, benefits, and issues associated with the advanced support tools provided during Mitigation. In particular, we focus on comments that gave clues to what aspects of the Mitigation configuration may have been associated with reducing workload as well as several related issues, such as why crew workload either remain flat or possibly decreased slightly with time delay in Mitigation, while flight controller workload increased by roughly the same amount as in Baseline.

Texting

Crewmember comments about texting indicate that the ability to communicate via texting greatly reduced crew/ground coordination difficulty (and possibly workload):

"Time delay made it difficult to do voice comm and still keep your place in procedures since the time is long enough the crew moves onto other tasks while waiting for the MCC to get back in touch for further direction. Texting was more effective than voice since we could go back and refer to the various recommendations from the ground."

"Ground or crew could go back and read IM [instant message] transmissions if they forgot or needed to reference for any reason."

According to these comments, texting reduced coordination difficulty because it freed crew from having to monitor the voice channels. Unlike Baseline, crew did not need to write anything down; they could concentrate on their work of the moment and refer to the texting log later, if need be.

Flight Controllers had many similar comments about texting (and WebPD, which will be discussed in the next section):

"Workload for [Control Center] CC was much lower by adding texting capability. My job switched from a listening mode to scanning/monitoring mode – watching the crew-progress in WebPD and watching for new texting messages."

In summary, the comments about the texting tool suggests that the reduction in crew workload at the longer delays in the Mitigation configuration was partly due to a reduced need for monitoring and communicating via voice loops, which freed up mental capacity to more effectively manage multitasking situations. However, FCT members provided additional comments indicating that the texting tool also *increased* workload:

"Workload was higher because with Text I had to monitor what conversations were on text and which ones were on audio."

"Two separate Text windows plus voice loops made more things to monitor."

We will have more to say about these "countervailing" comments in the following section on WebPD.

WebPD

WebPD was also enthusiastically received by crew and ground, and WebPD comments also shed important

additional light on the workload and coordination results. The following are two highly representative comments about the benefits of WebPD from FCT members:

"WebPD made it very easy to follow along in the procedures even with the time delay"

"Very easy to see where the crew should go from the line they were on as well as where they were going".

"The ability to track procedures and where the crew was in each step was awesome".

Not only did WebPD help the ground keep track of where the crew with within a procedure, but several mentions were made of the usefulness of the windows that showed what procedures were currently active, and which procedures had been completed:

"[I liked] [Ability to] see when crew brings up and starts a procedure, can see when they are done with a procedure."

In addition, many FCT members noted that the higher situation awareness afforded by WebPD enabled them to cut down on communications:

"WebPD helped because I didn't have to ask them about activities".

Finally, several comments hinted that texting and WebPD were complementary when it came to reducing workload:

"Workload for CC was much lower by adding [texting] Chat capability. My job switched from a listening mode to scanning/monitoring mode – watching the crew-progress in WebPD and watching for new chat messages."

These observations all pointed to a role for WebPD in reducing workload. However, just as with the texting tool, some comments from FCT members revealed another perspective:

"Workload was actually more noticeable because we actually had insight into the progress of the procedure from WebPD".

We have noted several times that ground has more of a responsibility for monitoring and maintaining high situation awareness of Crew activities than the other way around. Essentially, as the Mitigation tools increased the amount of information available to FCT, flight controllers spent more time understanding what the crew was doing by monitoring these sources, thus increasing their workload. This partially explains the different workload pattern between crew and ground in Mitigation, where the impacts of the tools were almost exclusively to reduce workload for the crew, but a mix of workload benefits and penalties for the ground.

ACAWS

Recall that the ACAWS technology provided two different forms of automated assistance with FDIR activities: Automated fault diagnosis, and automated recommendation of fault isolation or recovery procedures. Flight controller comments again indicate both workload reduction and a reduction in the need for coordination followed from these capabilities:

"ACAWS provided useful direction for the crew, so there was little need for us to do anything other than concur"

Similarly, crew comments were positive, and hinted that the tool allowed the crew to proceed more autonomously than in Baseline:

"ACAWS told me which procedure to work which the ground later confirmed but I had already completed the procedure."

The last quote speaks to both the situation awareness and autonomy issues, and also notes the benefits of greater autonomy for mitigating the effects of time delay:

"The time delay had little impact because ACAWS ran most of the procedure. Since the ground and crew can follow ACAWS, it was pretty seamless. MCC and DSH were able to come to common agreement with ACAWS. MCC and DSH statused each other via voice calls and texting."

8. COMMUNICATIONS ANALYSIS

The following comment is representative of a large number of similar comments to the effect that in the Mitigation configuration, the texting capability, the heightened situation awareness afforded by WebPD, and the more autonomous operational concept afforded by ACAWS all contributed to reducing the net amount of bi-directional communication required between the crew and the ground.

"Having ACAWS and Pidgin [texting] were useful in keeping track of crew activities and detailed information without having to call down to the crew".

Table 6. Voice only communications analysis.

Total talk time (s)			
Delay	Baseline	Mitigation	Ratio (M/B)
50	16158	12893	0.80
300	17060	11384	0.67

We evaluated this possibility by quantifying the total time that flight controllers and crew were engaged in communications in Baseline and Mitigation. Table 6 shows the total time spent by flight controllers and crew communicating over the voice channels. This includes time spent by flight controllers communicating with each other as well as all 'air-to-ground' communications. The times are accumulated separately for all runs at 50 seconds and 300

seconds of time delay. As shown in Table 6, there is no apparent difference in the amount of voice communication as a result of the time delay, either in the Baseline or Mitigation configurations. It is also apparent that less voice coordination takes place in the Mitigation configuration at both time delays; the final column shows the proportional reduction (ratio) of voice communication.

Table 7. Voice and texting communication analysis.

Total Voice + Chat Time						
	Baseline	Mitigation Chat	Mitigation Voice	Mitigation Total	Ratio (M/B)	Voice only (M/B)
50	16158	1702.5	12893	14595.5	0.90	0.80
300	17060	1856.5	11384	13240.5	0.78	0.67

Recall from Table 5 that there are four significant differences between the Baseline and Mitigation configurations: WebPD, ACAWS, Text, and the sharing of status between the crew and the flight control team through these tools that was not available in Baseline. One obvious possibility for the reduction in talk time from Baseline to Mitigation was simply that texting simply replaced voice as a means of communication, but the total volume of communication was unchanged. Table 7 combines Text and Voice communications in an attempt to determine whether the total volume of communication was reduced, or was just split between the two communication channels. In order to do so, the logged texting messages were assumed to be uttered at a rate of 2 words per second. The resulting time was added to the times spent using voice in Table 6. The results show that communication was still reduced during Mitigation when compared to Baseline configuration. It is also interesting to note how much communication was performed with texting instead of voice; roughly 11% of communications at 50 seconds time delay were performed through texting, versus roughly 13% of communications at 300 seconds.

9. Conclusions and Future Work

The Autonomous Mission Operations experiment provided a unique exploration of the reactions of NASA-trained flight controllers and astronauts to time delayed mission operations both in today's operations environment and in a environment featuring a suite of advanced automation tools and associated user interfaces. We conducted a series of runs of a two hour quiescent mission timeline. Each run was characterized by a time delay, ranging from 50 to 300 seconds one-way; a scenario, either Nominal, Systems failure, or Medical emergency; and a spacecraft and operations protocol configuration, either Baseline or Mitigation. Specifically, the Mitigation configuration:

- Included texting as a means of communication in addition to Voice.
- Added ACAWS to improve fault management.
- Replaced PDF versions of procedures with electronic procedures and the WebPD to track procedure execution and receive recommendations from ACAWS.
- Provided sharing of WebPD across the air-ground link.

Insight into the effects of time delay were gleaned through subjective questionnaires, evaluation of the number of completed activities, and analysis of voice and (in the case of the Mitigation configuration) texting logs. Our key findings are:

- Workload ratings and coordination difficulty between the flight control team and the crew *increased* with time delay.
- Workload and coordination difficulty *decreased* as a result of the mitigation configuration.
- Communications acts *decreased* in the Mitigation configuration.
- Flight controller workload ratings responded differently to configuration and time delay than the crew workload; specifically, crew workload was reduced by time delay in the Mitigation configuration, while flight controller workload increased with time delay regardless of configuration.
- Text, WebPD, and ACAWS were all explicitly identified as contributions to reduction in workload for the crew, but acted to both *reduce and increase* workload for the flight controllers.
- In contrast to these subjective and objective measures of performance, the actual number of tasks completed showed little effect of either time delay or configuration.

To begin to come to grips with all these findings and pull them all together, we begin with the results of a survey question completed by all participants at the end of the study:

"Taking into consideration all the scenarios, tasks, procedures, operational protocols, etc. that you experienced on this project, how would you rate the fidelity of the operations testing environment compared to an actual mission? (1 = Very low fidelity, 3 = Medium fidelity, 5 = Very high fidelity)"

The average over all participants was 3.1, or exactly in the middle of the scale, indicating that the experiment was performance in the context of a "medium fidelity" simulation. In their explanation for this rating, participants pointed over and over again to the fact that the DSH systems were not nearly as complex as the systems onboard a real spacecraft. Recall that the workload ratings were collected on the Bedford workload rating scale, which asked participants to choose heir rating against a "spare attentional capacity" dimension: In the run just completed, how much spare capacity would you have had to handle more tasks, should additional task requirements have been imposed? The fact that most ratings fell at or near the boundary between the low (Green) and intermediate (Yellow) ranges indicates

that we carried out our evaluation of time delay in a context where participants virtually always had capacity to spare. Two consequences follow from this supposition. First, spare capacity was available in even our most difficult workload situations. We suggest that participants utilized that spare capacity (casually speaking, they simply "tried harder") to ensure that they got their allocation of tasks completed, even in Baseline and even in the 300 sec time delay condition. This would account for why the empirical measures on task completion were barely affected by either configuration condition or time delay. Instead, the effects of these variables were manifest in terms of subjective workload. The second consequence follows directly from this observation by one of the participants:

"These sims were useful for testing new tools and comm delays with the crew - but not high enough fidelity for real procedure and execution tests. I suspect time delays in malfunction scenarios with far more complicated procedures would be far more challenging than we experienced in this lower fidelity environment."

The take-home message from this is simple: In a higher-fidelity simulation, participants would not have had (as much) spare capacity to handle the difficulties brought on by time delay, and so some of the impact would have been on task performance itself. In other words, the results of our experiment *underestimated* the impact of time delay on future deep-space mission operations, particularly for the most dangerous off-nominal situations during dynamic flight phases, when the most safety-critical and complex systems are all operational. Indeed, the fact that we did find clear impacts of time delay even in our less critical operational environment speaks to the ubiquity of time delay effects, and reinforces the need for ongoing work to develop, test, and validate next-generation operational concepts for next-generation missions.

Future work should fall into two different categories. The first category is further analysis of the data collected during this experiment. One example of further analysis is the assessment of data by scenario type; for instance, is there a relationship between the time delay, configuration, and scenarios? Another example is by flight controller position; for instance, is there a deeper relationship between individual flight controller roles (e.g. CapCom), time delay, and configuration? A third area of work is to analyze the voice and text logs to perform a deeper analysis of communications; were different words or phrases used with different frequency in different configurations? Were different words or phrases used in Text messages than were used in voice?

The second category of future work would be to conduct different experiments to learn more about the impact of time delay on operations. As noted in the comments about experiment fidelity, the spacecraft system and duration of runs impacted the fidelity. An experiment such as the ones conducted here should be performed in a higher fidelity environment in order to refine the lessons learned. In addition, additional experiments are needed to refine what elements of the mitigation configuration are truly responsible for influencing crew and flight controller workload and coordination difficulty. For example, it may be worthwhile to evaluate each of the four elements of the Mitigation configuration independently.

10. ACKNOWLEDGEMENTS

The Autonomous Mission Operations project would not have been possible without the participation of a large number of people. Timothy Hall of Johnson Space Center helped initiate this project. Jeffery Mauldin, Cynthia Sweitzer, Megan Rosenbaum, Lisa Voiles and Ernest Bell of Johnson Space Center served as our simulation coordinators. Katherine Garcia, Victor Hurst III, Chloe Abraszinskas and Ramon Gonzalez of Johnson Space Center served as subject matter experts to ensure the simulation was as realistic as it could be. Lindolfo Martinez, Tom Rich, Steve Rader, David Mongan and Jason Thompson of Johnson Space Center integrated the AMO technology into the Operations Technology Facility of the Christopher Kraft Mission Control Center. Marc Siebert, Michael Downs and Philip Michael of Kennedy Space Center handled communications, including the time delay software. Kerry McGuire of Johnson Space Center assisted with experiment design and ensured our questionnaires were complete and fielded during the experiments. Todd Milam and Deborah Schreckenghost of Johnson Space Center assisted with data analysis. Dennis Lawler, Larry Garner, Thomas Matthews, Arthur Molin and Matthew Hall of Johnson Space Center, and Gordon Aaseng, Baskaran Vijaykumar, Michael Dalal, Charles Fry, Chris Knight, May Windrem, Mei Wei, Jenessa Lin, Keith Swanson, Sriram Narasimhan, Jon Ossenfort and Charles Lee of NASA Ames Research Center all worked tirelessly to ensure that the Deep Space Habitat and the AMO software were in shape for the demonstration. Finally, we wish to thank the several dozen Johnson Space Center flight controllers and astronauts who served as our test subjects. This work was funded by the NASA Advanced Exploration Systems (AES) Program.

REFERENCES

[1] J.E. Bleacher, J.M. Hurtado, Jr., K.E. Young J. Rice, W.B. Garry, and D. Eppler. Desert Rats 2010 Operations Tests: Insights from the Geology Crew Members. Proceedings of the 42d Lunar and Planetary Sciences Conference, 2011.

[2] Brooke, J (1996). SUS: A "quick and dirty" usability scale. In Jordan, P.W., Thomas, B., Weerdmeester, B. A., & McClelland, A. L. (Eds). Usability Evaluation in Industry. Taylor & Francis, London, UK (pp. 189-194).

[3] Chappelle, S. Andrew F. Abercromby, Ph.D., Michael L. Gernhardt, Ph.D. NEEMO 15: Evaluation of Human Exploration Systems for Near Earth Asteroids. Proceedings of the Global Space Exploration Conference, 2012.

[4] Steven P. Chappell, Andrew F. Abercromby, William L. Todd, Michael L. Gernhardt. Final Report of NEEMO 14: Evaluation of a Space Exploration Vehicle, Cargo Lander, and Crew Lander during Simulated Partial-gravity Exploration and Construction Task. NASA Technical Report 2011-216152, 2011.

[5] Frank, J. When Plans are Executed by Mice and Men. Proceedings of the IEEE Aerospace conference, Big Sky, MT. 2010.

[6] Izygon, M., Kortenkamp, D., Molin, A., "A Procedure Integrated Development Environment for Future Spacecraft and Habitats," Space Technology and Applications International Forum, Albuquerque, NM, 2008.

[7] V. W. Hurst IV, S. Peterson, M.D. K. Garcia, A. Sargsyan, D. Ebert, D. Ham, B.S. D. Amponsah. S. Dulchavsky, Smart Ultrasound Remote Guidance Experiment (SURGE) –Ultrasound Image Collection during Lunar and Near Earth Orbit Space Missions. 82nd Annual Scientific Meeting of the Aerospace Medical Association; 8-12 Mayu 2011; Anchorage, AK; United States

[8] Kanas, N., Saylor, S., Harris, M., Neylan, T., Boyd, J., Weiss, D., Baskin, P., Cook, C., Marmar, C. High vs. Low Crewmember Autonomy in Space Simulation Environments. Acta Astronautica, v. 67, no. 7-8, 2010 p. 731 - 738

[9] Kanas, N., Harris, M., Neylan, T., Boyd, J., Weiss, D., Cook, C., Saylor, S. High vs. Low Crewmember Autonomy during a 105-day Mars Simulation Mission. Acta Astronautica, v. 69, no. 7-8, 2011 p. 240 – 244

[10] Kennedy, Kriss J. NASA Habitat Demonstration Unit Project – Deep Space Habitat Overview. 41st International Conference on Environmental Systems (ICES), Portland, Oregon, USA, 17-21 July 2011.

[11] Kortenkamp, D., Verma, V., Dalal, K.M., Bonasso, R.P., Schreckenghost, D., Wang, L., "A Procedure Representation Language for Human Space Flight Operations" 9th International Symposium on Artificial Intelligence, Robotics, and Automation for Space, Los Angeles, CA, 2008.

[12] Kortenkamp, D., R.P., Schreckenghost, D., Wang, L., "Developing and Executing Goal-Based, Adjustably Autonomous Procedures" AIAA InfoTech@Aerospace Conference, AIAA, Ronhert Park, CA 2007

[13] Mackey, R., Brownston, L., Castle, J., Sweet, A. Getting Diagnostic Reasoning Off the Ground: Maturing Technology With TacSAT3. IEEE Intelligent Systems, Sept. Oct. 2010 p. 27 – 35.

[14] A. Mathur, S. Deb, and K. Pattipati, "Modeling and Real-Time Diagnostics in TEAMS-RT," Proc. American Control Conf., IEEE Press, 1998, pp. 1610–1614.

[15] McCandless, J. W., McCann, R. S., Berumen, K. W., Gauvain, S. S., Palmer, V. J., Stahl, W. D., & Hamilton, A. S (2005). Evaluation of the Space Shuttle Cockpit Avionics Upgrade (CAU) Displays. In Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society.

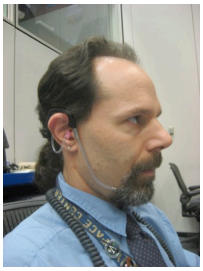
[16] McCann, R. S., Beutter, B. R., Matessa, M., McCandless, J. W., Spirkovska, L., Liston, D., Hayashi, M., Huember, V., Lachter, J., Ravinder, U., Elkins, S., Renema, F., Lawrence, R., & Hamilton, A. (2006). Evaluation of an Onboard Real-Time Fault Management Support System for Next-Generation Space Vehicles. Report to the Human Research Program

[17] Schwabacher, M., Martin, R., Waterman, R., Oostdyk, R., Ossenfort, J., and Matthews, B. Ares I-X Ground Diagnostic Prototype. AIAA Infotech@Aerospace Conference, Atlanta, April 2010.

[18] Tri, T., Kennedy, K., Toups, L., Gill, T., Howe, A. S. Planning and Logistics for the In-Field Demonstration of NASA's Habitat Demonstration Unit (HDU) Pressurized Excursion Module (PEM) at Desert Rats 2010. Proceedings of the International Conference on Environmental Systems, Portland OR. 2011.

BIOGRAPHIES

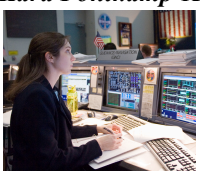
Jeremy Frank Dr. Jeremy Frank works in the Autonomous Systems and Robotics Area of the Intelligent Systems Division at NASA Ames Research Center. He is Group Lead of the Planning and Scheduling Group, an organization of 30 researchers and staff that designs and builds space mission operations tools. Dr. Frank has an MS and a Ph. D. in Computer Science from the University of California (Davis). He also has a B.A. in Mathematics from Pomona College. Dr. Frank's research interests include combinatorial optimization, operations research, artificial intelligence, applied to the automation of space mission operations. Dr. Frank has successfully led and participated in several research and development projects during his 12 years at NASA. Presently, Dr. Frank is the Principal Investigator of the Autonomous Mission Operations (AMO) Project, a NASA Advanced Exploration Systems Program project that develops advanced technology prototypes for mission operations in the presence of large time delays. He has received over twenty NASA awards, including a Silver Snoopy (the Astronaut's Personal Award).



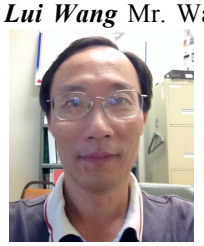
Rob McCann Robert S. McCann holds a Ph. D. in Cognitive Psychology from the University of Waterloo. Following the completion of his doctorate, he was awarded a National Sciences Foundation Postdoctoral Fellowship to conduct basic research in human multi-tasking performance limitations at NASA's Ames Research Center. In 2000, Dr. McCann was hired into the Information Sciences and Technology Directorate as a NASA research psychologist. His responsibilities included managing and planning Ames efforts to redesign the shuttle cockpit display formats to enhance crews' situation awareness and reduce workload. His most recent position has been Group Lead of the Intelligent Spacecraft Interface Systems Laboratory, which Dr. McCann developed to design and evaluate cockpit displays and user interfaces for the new generation of spacecraft that will take us back to the Moon and beyond. He is the author or co-author of over 60 technical papers and research publications, which have been cited in over 700 articles in the open scientific literature.

Lilijana Spirkovska Dr. Lilly Spirkovska is a Computer Engineer at NASA Ames Research Center working in the area of Integrated System Health Management (ISHM with a special interest in human interaction with ISHM systems – what information to present and how to present that information to make the operator most effective in managing the health of systems. Lilly has a Ph. D. in Computer Science, an MS in Computer Science, a B.S. in Computer Science and Mathematics, and is a general aviation commercial, instrument-rated pilot. She has been with NASA for over 26 years working in diverse areas such as neural networks for pattern recognition, automation systems for improving air traffic control both in the air and on the airport surface, aviation weather visualization and automation systems for general aviation pilots, and, most recently, ISHM systems for eventual use by astronauts and controllers.

Kara Pohlkamp Kara Pohlkamp is an employee of NASA at Johnson Space Center. She is a Guidance, Navigation and Control flight controller who sat console in Mission Control for the Space Shuttle and now leads operations work for MPCV/Orion. She also is the JSC lead for the Autonomous Mission Operations (AMO) Project. Kara holds a M.S. in Aerospace Engineering from the Georgia Institute of Technology (2007) and a B.S. in Aerospace Engineering and Public Relations from Iowa State University (2005).



Lui Wang Mr. Wang has worked at NASA Johnson Space Center since 1983. For the past 28 years, he has led the research and infused numerous automation and intelligent software technologies into the Space Shuttle, International Space Station, and the Exploration Programs for the Software Robotics and Simulation Division. He holds 2 US Software Patents on Intelligent Computer Aided Training and Dynamics Pattern Matching using Incomplete Data. Mr. Wang led the development Constellation Electronic Procedure System for the NASA Constellation Program. He is collaborating with NASA Ames Research Center in maturing electronic procedure technology for near term NASA human spaceflight programs. Also, he is co-leading the development of the avionics and flight software the Advanced Exploration Systems (AES)/Habitation, Autonomous Mission Operations Projects and NASA Office of Chief Technology (OCT) / Autonomous System projects.



Lee Morin Lee Morin M.D. Ph.D. is a captain in the US Navy and a NASA astronaut mission specialist. Lee flew aboard Atlantis on shuttle mission STS-110. He performed two EVAs to install the S0 truss on the International Space Station. Since 2005 Lee has been leading the Astronaut Office Rapid Prototyping Lab, which has been leading development of the displays and controls for the Orion Multi-Purpose Crew Vehicle (MPCV) spacecraft.