Promoting Crew Autonomy in a Human Spaceflight Earth Analog Mission through Self-Scheduling

Jessica J. Marquez[1]
NASA Ames Research Center, Moffett Field, CA, 94035, USA

Shivang Shelat[2]
San Jose State University, San Jose, CA, 95112, USA

John A. Karasinski[3]
NASA Ames Research Center, Moffett Field, CA, 94035, USA

Deep space exploration missions face the challenge of communication transmission latencies between ground stations and astronaut crews due to increasing distance between the Earth and spacecraft in transit. To address this, research at NASA has aimed toward supporting crew autonomy by enabling astronauts to schedule their own timelines with minimal oversight from Mission Control. While self-scheduling has been shown to be feasible, it is yet to be studied as an integral part of autonomous crew operations. The current paper reviews the operationalization of self-scheduling and a number of related objectives during Campaign 6 of HERA, a Human Exploration Research Analog. Research objectives include studying the effects of phasic autonomy over the course of a 45-day mission, evaluating differences in scheduling performance produced by software interface aids, and deploying a novel measure of crew attitudes toward self-scheduling and plan execution.

I. Nomenclature

LDEM = Long Duration Exploration Mission
HCAAM = Human Capabilities Assessments for Autonomous Missions
VNSCOR = Virtual NASA Specialized Center of Research
ISS = International Space Station
NEEMO = NASA Extreme Environment Mission Operations
HERA = Human Exploration Research Analog
C6 = Campaign 6
JSC = Johnson Space Center
HRP = Human Research Program
MC = Mission Control
MD = Mission Day
TLX = Task Load Index
n = number of subjects

---

[1] Human-Systems Researcher, Human-Computer Interaction Group, NASA Ames Research Center, AIAA Member.
[2] Student Researcher, Human-Computer Interaction Group, NASA Ames Research Center.
[3] Research AST, Human/Machine Systems, Human-Computer Interaction Group, NASA Ames Research Center, AIAA Member.

II. Introduction

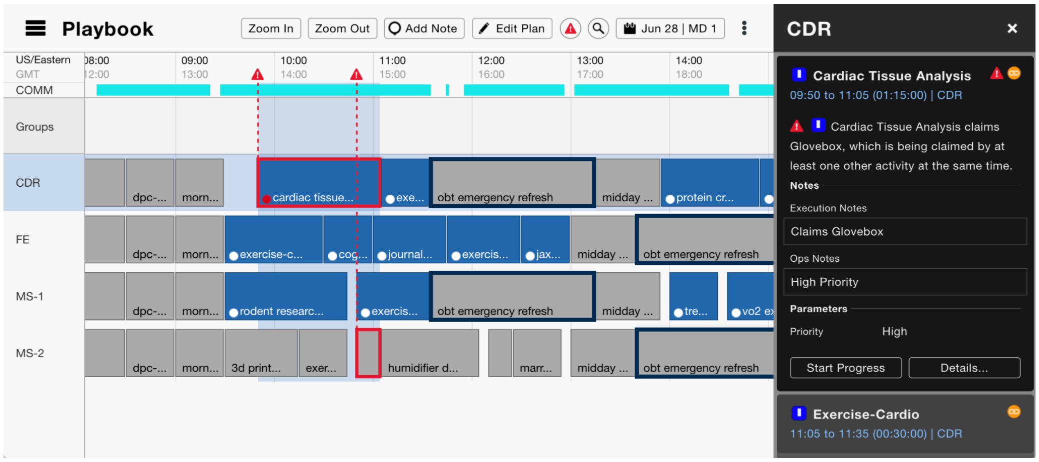
NASA is actively planning to conduct long duration exploration missions (LDEMs) that require astronauts to operate more autonomously from flight controllers on Earth. To enable crew autonomy, it is necessary to define a new concept of operations and to research the different aspects of autonomy required for mission success. To this end, NASA funded the Human Capabilities Assessments for Autonomous Missions (HCAAM) Virtual NASA Specialized Center of Research (VNSCOR). This VNSCOR includes seven research projects funded across government and academia and is focused on promoting crew autonomy. Our VNSCOR project is focused on enabling crew self-scheduling through human-centered software development and operational research.

Future LDEMs will delve deeper into space and face a number of challenges, among the most critical of which is increased communication transmission latencies between flight controllers and crews. Under the current concept of operations, these latencies will result in significant idle time as crews wait for flight controller response. Enabling astronauts to schedule or reschedule their own timeline may reduce crew idle time and improve crew autonomy, but research is needed to enable crewmembers to schedule effectively.

A. Playbook – A Self-Scheduling Tool

The Human-Computer Interaction Group at NASA Ames Research Center has spent the past decade developing Playbook (see Fig. 1), a mission planning and scheduling software tool which allows novice planners to organize activities and create feasible timelines [1]. Timelines must be created in accordance with a variety of constraints and a limited availability of resources – for example, a science activity may require a particular crew member, need to done at a certain time of day, and/or needs to be done using a particular resource (i.e., the only glovebox available). Playbook's display highlights any violations of these constraints, prompting the user to reschedule tasks until the timeline is valid and free from constraint violations. The integration of such interface aids streamlines the timeline creation process [2].

Our previous research has characterized human performance in a number of studies using Playbook [3, 4] in an effort to identify what are the key elements that drive self-scheduling complexity. Furthermore, we have demonstrated that self-scheduling is a feasible undertaking for astronauts in spaceflight-relevant environments such as the International Space Station (ISS) [5] and NASA analogs [1]. Human spaceflight analogs are environments that mimic some aspect of spaceflight and are used to evaluate mission operations in an ecologically valid context. Over the past several years, we have provided opportunities to over sixty analog crew members to conduct self-scheduling in the NASA Extreme Environment Mission Operations (NEEMO) [1] and Human Exploration Research Analog (HERA) analogs [6]. During a Technology Demonstration onboard the ISS, an astronaut scheduled an afternoon during spaceflight operations [5]. Similarly, participants in the aforementioned analogs scheduled activities over a short period of time (e.g., an afternoon) or rescheduled a few activities in their timeline.



Cardiac Tissue Analysis
09:50 to 11:05 (01:15:00) | CDR

Cardiac Tissue Analysis claims Glovebox, which is being claimed by at least one other activity at the same time.

Notes

Execution Notes: Claims Glovebox

Ops Notes: High Priority

Parameters

Priority: High

Exercise-Cardio
11:05 to 11:35 (00:30:00) | CDR

Fig. 1 Playbook, a mission planning and scheduling software tool designed to support crew autonomy. A violation of timeline constraints is highlighted in the display.

Up until now, we have focused our self-scheduling research on individual trials and events outside of crews' operational timeline. Looking forward, it becomes critical to evaluate crew self-scheduling as a core element of autonomous space operations. HERA Campaign 6 (C6), which started in October 2021, will contribute to this goal by informing a number of research questions regarding crew self-scheduling. It utilizes a novel experimental paradigm, providing crews an opportunity to directly schedule their operational timeline.

HERA is an isolation analog located at NASA Johnson Space Center (JSC) that simulates future LDEMs. The analog's broad objective is to study crew health, performance, and autonomy in a confined environment [7]. In regards to self-scheduling, C6 has three primary aims; first, its missions will simulate shifts in autonomy that are present during actual spaceflight operations. As spacecraft travel farther and farther from the Earth, communication delays grow and the necessity of crew autonomy increases. This dynamic will be simulated across all crew communications, allowing us to evaluate self-scheduling behaviors in an untested context. Second, crewmembers in C6 will schedule using new Playbook features. For half of C6's missions, aids that were designed to simplify the self-scheduling process will be integrated into the tool's interface. We will measure differences in scheduling performance, workload, and usability produced by the designs between the missions. Testing aids in an analog environment such as HERA may support their adoption for actual operations in future missions. Finally, we will deploy a measure of "plan goodness" to bridge the gap between self-scheduling decisions and the perceived quality of the plan that was created. Timelines are executed by the crew as part of their nominal operations, allowing us to evaluate planner strategies, preferences, and attitudes associated with the plans they create. This paper will review C6's experimental design and assess these objectives related to self-scheduling and crew autonomy.

III. HERA Campaign 6

For each of the four missions that make up HERA C6, NASA selects four analog crew members who have similar profiles to the astronaut population. These crew members continuously live in the habitat for 45 days and only interact with their family, friends, and mission control remotely. Crews perform scientific and operationally-relevant activities, typically consisting of research tasks defined by the Human Research Program (HRP). A Mission Control (MC) supports the completion of the activities, similar to the ISS Mission Control Center. Communication transmission delays are inserted between MC and crew; for C6, the one-way transmission latency ranges from 30 seconds to 5 minutes. At the end of each mission, there is a comprehensive debrief during which researchers can ask a number of questions to the crew to gain further insight across the various research tasks completed during the mission.

A. Operationalizing Self-Scheduling

HERA C6 mission managers decided to operationalize crew autonomy, in part, as the ability to conduct self-scheduling. As such, self-scheduling is an option for most of the mission for the first time in a NASA spaceflight analog. Thus, C6 presents a unique opportunity to study the effects of operationalizing self-scheduling as an integral part of autonomous crew operations. HERA's initial mission timeline is created by the MC Planner. They are responsible for scheduling all the crew's required activities, which are driven by the scientific and operational objectives of the Campaign. The HERA MC Planner understands all the constraints that exist around the various required activities. Example constraints include that certain activities must be done by a predetermined time, or that a particular experiment must be completed by all crew members on certain days. The MC Planner also knows what resources are available – for instance, there is only one exercise machine, or a specific experiment requires the coordination of two crew that must use the only table. All of the mission constraints and resources are integrated into a feasible timeline. For HERA C6, the MC Planner coordinated with lead researchers to determine which scientific and operational activities were *flexible*. Flexibility is an indicator to the crew that an activity can be rescheduled, and flexible activities were pre-determined by the mission. Some activities are not flexible because they require significant coordination between crew and MC, are specifically required to be at a particular time of day, or have other scientifically-driven constraints.

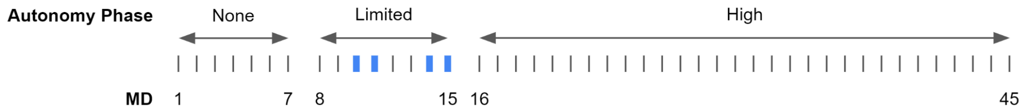


Fig. 2 The three different autonomy phases throughout the 45 Mission Days (MDs), with our four "Autonomous Days" highlighted in blue.

For C6, each 45-day mission is divided into three phases of autonomy which determine the crew's ability to reschedule their activities (see Fig. 2). In the first phase (No Autonomy), which lasts a week, the crew's timeline is fixed and no activities can be rescheduled. In the second phase (Limited Autonomy), which is eight days long, the crew's timeline has a significant portion of activities marked as flexible. During this second phase, any rescheduling completed by the crew must be approved by the MC Planner. In the third phase (High Autonomy), which is thirty days, self-scheduling is entirely unprompted and only initiated by the crew themselves. In this phase, the MC Planner only makes changes when the self-scheduling causes violations in the resulting timeline.

Our controlled research experiment takes place during Limited Autonomy. During this time, we specifically requested that the crew discuss their scheduling preferences and to self-schedule most of an operational day. In turn, this resulted in four preference meetings and four autonomous (or self-scheduled) days. Preference meetings are allocated blocks of time for one of the crew members, assigned as the planner, to lead a team discussion on what their scheduling preferences may be. On the autonomous day, which takes place after the preference meeting day, the crew planner uses Playbook to actually schedule a timeline for the whole crew. A majority of the activities for that timeline are presented in a Task List, and the crew planner is required to schedule all of them for their assigned day. The MC Planner then reviews and approves the day after the crew planner is finished scheduling. The autonomous days are the only times during the mission when the crew is explicitly required to schedule activities. While attempts were made to make these autonomous days similar, they varied greatly in terms of the number of activities, total duration of activities, and the number of constraints present on and between activities (see Fig. 3). This reflects the nature of astronauts' operational timelines, where crewmembers are conducting a variety of science and maintenance tasks, and few days are the same.

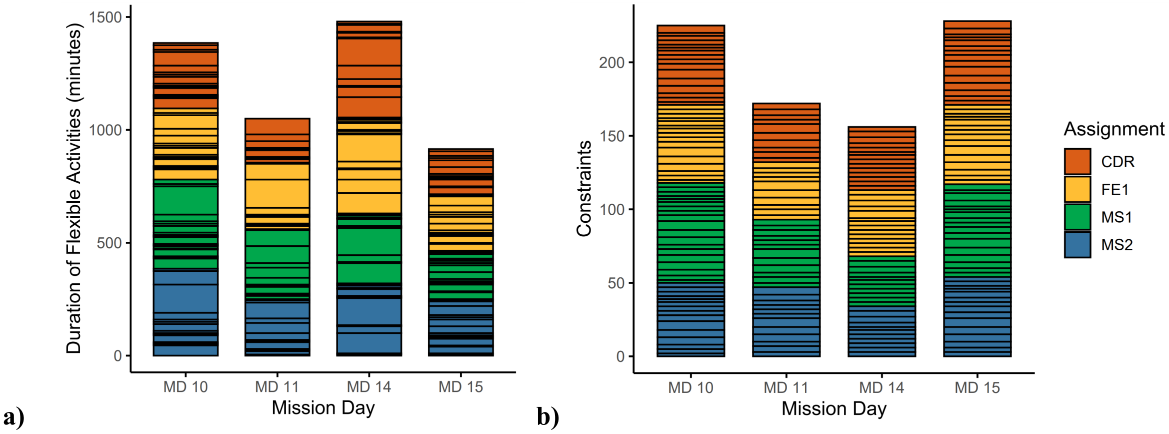


Fig. 3 A visualization of the differences in total number of flexible minutes and constraints for each Autonomous Day. The colors indicate which crew member a particular activity was assigned to.

A variety of data is collected to evaluate self-scheduling performance during these autonomous days. Primarily, we are able to study interactions within Playbook from the tool's data logs, such as self-scheduling efficiency (i.e., time on task) and effectiveness (i.e., number of violations created) [8]. After they finish self-scheduling, the planner fills out the NASA-Task Load Index (TLX) [9], a standard measure of subjective workload. During both preference meetings and the autonomous days, crew members wore voice recording devices, allowing for post-hoc observation and insight into the nature of their collaboration, strategies, and preferences. At least two days after the autonomous day, the crew executes the created plan. At the end of this execution day, the planner fills out a measure of plan goodness, which evaluates attitudes toward the plan and the self-scheduling process. The development, content, and goals of this measure will be discussed in a subsequent section.

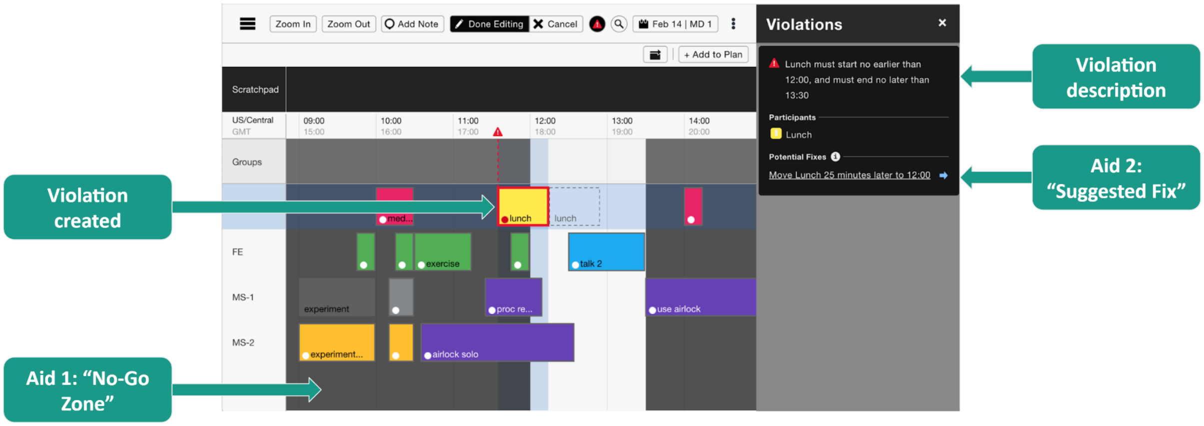
The third phase, High Autonomy, comprises the rest of the mission: thirty days in which the crew's timeline has a significant portion of flexible activities that can be rescheduled if desired and no approval is required by the MC Planner. We are able to see when and how often the crew takes it upon themselves to self-schedule. Coupled with debriefs at the end of the mission, we can also infer as to *why* this self-scheduling is occurring. At the end of this phase, crew members fill out two validated usability questionnaires in regards to Playbook: the System Usability Scale [10], which is recommended by NASA's *Human Integration Design Handbook* [11] to measure a system's usability, and the User Experience Questionnaire [12], which has been used to evaluate the tool's user experience in the past [13].

The three-phase scheduling paradigm effectively integrates the shifting nature of autonomy demands into experimental human spaceflight operations. We are able to study the effects of this dynamic on performance, workload, and general operations. However, C6's paradigm also allows us to test differences between missions – particularly, assistive Playbook features which are present for some crews but not for others.

B. Testing Scheduling Interface Aids

As part of our controlled research experiment in HERA C6, we aim to validate software aids designed to assist novice planners conduct self-scheduling. Our team has developed two visual aids (or countermeasures) for HERA: No-Go Zones and Suggested Fixes. We hypothesize that these countermeasures will facilitate self-scheduling performance. Specifically, we expect shorter self-scheduling times, fewer violations created while self-scheduling, and lower subjective workload ratings. Additionally, crew will be more likely to conduct self-scheduling throughout the mission. For HERA C6, the first two missions will not have access to the countermeasure aids while doing self-scheduling while the latter two missions will always have the aids enabled.

As mentioned earlier, each mission has a complex set of constraints and available resources that must be integrated into the timeline. A schedule is only feasible if all the constraints are met. These constraints are modeled as part of the activity by the MC Planner. Once the activity is scheduled, Playbook instantaneously checks if the activity meets its modeled constraints. For instance, the activity must be completed by a specific crewmember or two activities must be scheduled in a particular order. When an activity's constraints are not met, Playbook indicates a violation. The violation markers are distinct: red coloring and warning indicator. The countermeasure aids assist novice planners in two manners: provide more visual information about the constraints and provide self-scheduling solutions to activities in violation. The No-Go Zones (see Fig. 4) depict where *not* to schedule an activity on the timeline. The planner is briefly trained to understand that a violation will be created if they choose to schedule an activity within a No-Go Zone. We also have a second aid that suggests a self-scheduling fix (see Fig. 4). This secondary aid is included as some scheduling problems require multiple activities to be rescheduled, which cannot be illustrated with a single No-Go Zone visualization.



Lunch must start no earlier than 12:00, and must end no later than 13:30

Participants

Lunch

Potential Fixes

Move Lunch 25 minutes later to 12:00

Fig. 4 The Playbook scheduling aids present during Missions 3 and 4 in C6.

C. Evaluating Plan Goodness

In collaboration with research psychologists at NASA, we created a novel measure for plan goodness for C6. The goal of the questionnaire was to have the means of collecting subjective data on the quality of a self-scheduled, executed plan, as well as to get additional information on the self-scheduling task itself. Currently, we have objective metrics that describe the validity of a plan (e.g., no remaining violations, all required activities scheduled), but no self-report measure that captures individual attitudes toward planning and execution. Individual preferences and biases (e.g., "I'm a morning person") influence the perceived quality of a schedule, particularly if those creating the schedule also have to execute them. Furthermore, based on previous interviews with expert spaceflight planners, we learned that what makes a plan "good" varies from person to person.

The developed measure consists of eleven statements rated on a 100-point scale, with a high rating indicating that the subject strongly agrees with the statement. Afterwards, subjects respond to additional open-ended prompts to provide qualitative insight into preferences for a good plan, contributors to workload, attitudes toward the execution day, unexpected obstacles, and Playbook features that would have been helpful. The questionnaire is as follows:

1. *Creating the self-scheduled plan was cumbersome.*
2. *I created a good plan for our team.*
3. *I prefer to follow a prescribed timeline.*
4. *I was able to create a high-quality plan for today.*
5. *My team was more motivated to complete tasks today because of my self-scheduled plan.*
6. *Self-scheduling allowed me to complete more work today.*
7. *Self-scheduling allowed me to effectively manage my workload.*
8. *Self-scheduling increased my motivation to complete tasks today.*
9. *Self-scheduling provided me with a sense of control.*
10. *Today's self-scheduled plan was difficult to create.*
11. *Today's self-scheduled plan was difficult to execute.*

HERA C6 is the first experiment in which this questionnaire has been deployed. It allows us to characterize the crews' attitudes toward the self-scheduling task itself and the executed day that they plan. Furthermore, we can explore correlations between the statement ratings and current objective measures, such as workload, violations created, and time on task.

IV. Conclusion

A. Preliminary Results

HERA C6 is scheduled to be completed in October 2022. Here, we present some preliminary findings from Missions 1 and 2. Subjective workload, measured via the NASA-TLX questionnaire after crew members scheduled on the autonomous days, had an average of 39.1 and a standard deviation of 12.1 ($n = 8$). Notably, this number is comparable to merged workload scores from the scheduling condition of our controlled lab experiment [4], which had an average of 49.2 and a standard deviation of 19.1 ($n = 15$). While the NASA-TLX scores are slightly lower in the analog setting, the reported subjective workload seems to be approximately equivalent (although a complete dataset and a more direct comparison is needed to confirm this).

For the plan goodness questionnaire, a high rating indicates that the subject strongly agrees with that particular item (see Fig. 5). Results show that there is considerable variability in ratings for particular items between the missions. For example, Mission 1's crew rated statements 5 and 6 highly, suggesting that they see more benefits from self-scheduling than Mission 2's crew; however, Mission 1's crew seems to prefer a prescribed timeline.

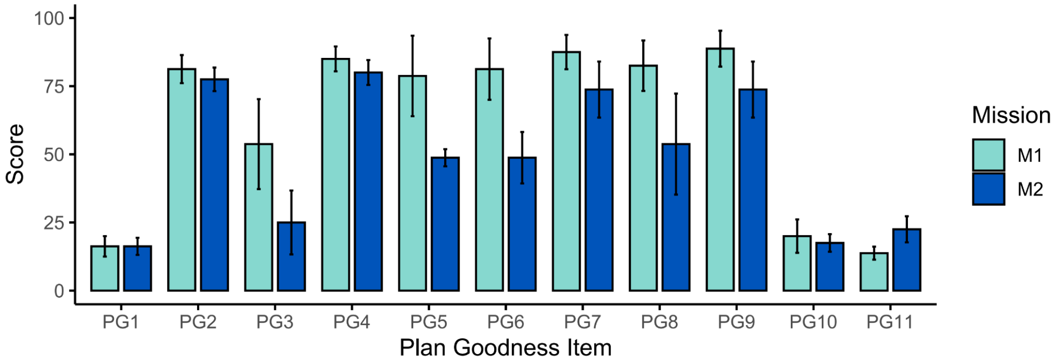


Fig. 5 A distribution of scores per Mission for each item in the plan goodness questionnaire. Error bars represent standard error.

Figure 6 presents the number of self-scheduling sessions occurring per MD. While completely optional (except for one session on each of the four autonomous days), self-scheduling is willingly conducted by HERA crew members. There appears to be an increase in self-scheduling as the crew enters the High Autonomy phase of the mission; however, there is high variability over days which remains unexplained. Mission 2's crew were more likely to conduct self-scheduling than Mission 1, despite Mission 1's crew seemingly having a more positive view of self-scheduling.

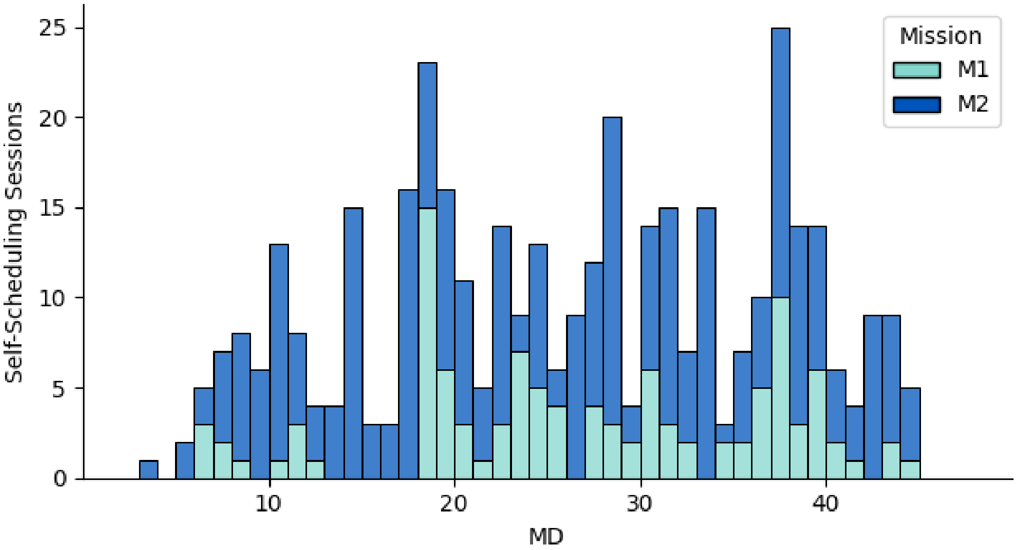


Fig. 6 The frequency of self-scheduling sessions over the course of each mission.

B. Future Work

Data collection is expected to be completed once C6 is over. We will conduct data post-processing of self-scheduling performance for the autonomous days, including determining time on task and number of violations, as well as assessing if the different characteristics (total duration of flexible activities, number of constraints, etc.) of each day impacted performance. Along with workload and usability, we will compare the effect of the countermeasure aids on performance (i.e., compare M1 and M2 against M3 and M4). Additionally, we plan on exploring relationships between the number of a day's self-scheduling instances and factors such as plan goodness statement ratings, phase transitions, timeline complexity, and social cohesion within the crew. In Fig. 6, there appears to be high variability in the frequency of these sessions – these fluctuations might be explained by a number of other variables which have yet to be explored.

Voice recordings are also being collected and must be post-processed, transcribed, and encoded for analysis. Recent trends in usability research move toward evaluating interactive tools in the context of where they will be used. Clinkenbeard [14] outlines an approach in which audio recordings during computer-supported cooperative work are leveraged to better understand how a tool will be used in more complex contexts. Up until now, our team has only evaluated Playbook's usability after controlled laboratory experiments and during analog debrief sessions; however, relying solely on this approach can obscure the details of human-machine interactions. In actual operations, Playbook's users will be in the confined, social environment of a spacecraft, which is effectively simulated by the HERA analog. C6's recordings provide real-time voice data from an operational environment, allowing us to characterize the nature of social interactions and collaboration during scheduling. For example, we may pose the question: are crew members collaborating to navigate Playbook's interface during the planning process, and is there a particular feature that they verbalize sentiment towards? Coupled with conventional usability scores from the questionnaires at the end of each mission, we can also infer whether Playbook is usable in the collaborative, complex environment where it was used.

Acknowledgements

This research was funded by the NASA Human Research Program's Human Factors and Behavior Performance Element (NASA Program Announcement number 80JSC017N0001-BPBA) Human Capabilities Assessment for Autonomous Missions (HCAAM) Virtual NASA Specialized Center of Research (VNSCOR) effort. The authors would like to thank Nick Bathurst for his comments in developing this manuscript. We would like to thank HERA planner Heath Clark as well as California State University, Long Beach interns Jessica Mar and Joshua Mosonyi for their inputs and contributions to this research.

References

[1] Marquez, J. J., Hillenius, S., Kanefsky, B., Zheng, J., Deliz, I., and Reagan, M., “Increasing crew autonomy for long duration exploration missions: Self-scheduling.” *2017 IEEE Aerospace Conference*, 2017, 1–10.
doi: 10.1109/AERO.2017.7943838

[2] Marquez, J. J., Pyrzak, G., Hashemi, S., McMillin, K., and Medwid, J., “Supporting Real-Time Operations and Execution through Timeline and Scheduling Aids.” *43rd International Conference on Environmental Systems*, Vail, CO., 14 July 2013.
doi: 10.2514/6.2013-3519

[3] Lee, C., Marquez, J., and Edwards, T., “Crew Autonomy through Self-Scheduling: Scheduling Performance Pilot Study.” *AIAA SciTech 2021 Forum*, VIRTUAL EVENT, 11 Jan. 2021.
doi: 10.2514/6.2021-1578

[4] Marquez, J. J., Edwards, T., Karasinski, J. A., Lee, C. N., Shyr, M. C., Miller, C. L., and Brandt, S. L., “Human Performance of Novice Schedulers for Complex Spaceflight Operations Timelines.” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2021.
doi: 10.1177/00187208211058913

[5] Marquez, J. J., Hillenius, S., Healy, M., and Silva-Martinez, J., “Lessons learned from International Space Station crew autonomous scheduling test.” *International Workshop on Planning and Scheduling for Space*, July 2019.

[6] Gale, J. W., Yashar, M., Karasinski, J., and Marquez, J., “Evaluation of Self-Scheduling Exercises Completed by Analog Crewmembers in NASA’s Human Exploration Research Analog (HERA).” *ASCEND 2021*, Las Vegas, Nevada & Virtual, 15 Nov. 2021.
doi: 10.2514/6.2021-4076

[7] LSDA Mission – HERA Campaign 6. URL: https://lsda.jsc.nasa.gov/Mission/miss/3421 [retrieved 8 August 2022].

[8] Kanefsky, B., Zheng, J., Deliz, I., Marquez, J. J., and Hillenius, S., “Playbook Data Analysis Tool: Collecting Interaction Data from Extremely Remote Users.” In *Advances in Usability and User Experience* (T. Ahram and C. Falcão, eds.), Springer International Publishing, Cham, 2018, pp. 303–313.
doi: 10.1007/978-3-319-60492-3_29

[9] Hart, S. G., and Staveland, L. E., “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research.” In *Advances in Psychology*, Elsevier, 1988, pp. 139–183.
doi: 10.1016/S0166-4115(08)62386-9

[10] Jordan, P. W., Thomas, B., McClelland, I. L., and Weerdmeester, B., “Usability evaluation in industry.” In *NASA: Human Integration Design Handbook (HIDH)*, CRC Press, NASA, 2010.

[11] *Human Integration Design Handbook (HIDH)*, NASA, 2010.

[12] Schrepp, M., Hinderks, A., and Thomaschewski, J., “Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios.” In *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience* (A. Marcus, ed.), Springer International Publishing, Cham, 2014, pp. 383–392.
doi: 10.1007/978-3-319-07668-3_37

[13] Shelat, S., Karasinski, J. A., Flynn-Evans, E. E., and Marquez, J. J., “Evaluation of User Experience of Self-scheduling Software for Astronauts: Defining a Satisfaction Baseline.” In *Engineering Psychology and Cognitive Ergonomics* (D. Harris and W.-C. Li, eds.), Springer International Publishing, Cham, 2022, pp. 433–445.
doi: 10.1007/978-3-031-06086-1_34

[14] Clinkenbeard, M., "Multimodal conversation analysis and usability studies: exploring human-technology interactions in multiparty contexts." *Communication Design Quarterly*, Vol. 6, No. 2, 2018, pp. 103–113.
doi: 10.1145/3282665.3282675