Human Performance of Novice Schedulers for Complex Spaceflight Operations Timelines

Jessica J. Marquez[1], Tamsyn Edwards[2], John A. Karasinski[1], Candice N. Lee[2], Megan C. Shyr[1], Casey L. Miller[2] and Summer L. Brandt[1], [1]NASA Ames Research Center, Moffett Field, CA, USA, [2]San Jose State University, San Jose, CA, USA

Objective: Investigate the effects of scheduling task complexity on human performance for novice schedulers creating spaceflight timelines.

Background: Future astronauts will be expected to self-schedule, yet will not be experts in creating timelines that meet the complex constraints inherent to spaceflight operations.

Method: Conducted a within-subjects experiment to evaluate scheduling task performance in terms of scheduling efficiency, effectiveness, workload, and situation awareness while manipulating scheduling task complexity according to the number of constraints and type of constraints.

Results: Each participant (n = 15) completed a set of scheduling problems. Results showed main effects of the number of constraints and type of constraint on efficiency, effectiveness, and workload. Significant interactions were observed in situation awareness and workload for certain types of constraints. Results also suggest that a lower number of constraints may be manageable by novice schedulers when compared to scheduling activities without constraints.

Conclusion: Results suggest that novice schedulers' performance decreases with a high number of constraints, and future scheduling aids may need to target a specific type of constraint.

Application: Knowledge on the effect of scheduling task complexity will help design scheduling systems that will enable self-scheduling for future astronauts. It will also inform other domains that conduct complex scheduling, such as nursing and manufacturing.

Keywords: scheduling, computer-supported collaborations, analysis and evaluation, multivariate analysis

Précis: Evaluated novice schedulers' performance as a function of number of constraints and different types of constraints pertinent to spaceflight operations. Differences detected in efficiency, effectiveness, workload, and situation awareness will help design future scheduling systems and aids.

Address correspondence to Jessica J. Marquez, NASA Ames Research Center, Mail Stop 262-2, Bldg. 262, Rm. 132, P.O. Box 1, Moffett Field, CA 94035-1000, USA; e-mail: jessica.j.marquez@nasa.gov

HUMAN FACTORS
Vol. 0, No. 0, ■■ ■, pp. 1-16
DOI:10.1177/00187208211058913
Article reuse guidelines: sagepub.com/journals-rmissions

INTRODUCTION

Planning and scheduling the activities required to operate the International Space Station (ISS) takes a team of highly trained and experienced planners weeks to complete (Dempsey, 2018). The task of planning and scheduling is complex because there are a myriad of requirements, resources, and constraints that must be satisfied. Failure to do so leads to schedules that are infeasible and potentially unsafe, leading to loss of mission objectives or threatening the health and safety of astronauts. Furthermore, there are often too many activities and insufficient time for astronauts to complete them all. Thus, ISS planners leverage sophisticated software tools to manage all the constraints that dictate when activities can be scheduled for astronauts. The successful timelines produced govern an astronaut's day-to-day life down to the minute and they must adhere to this schedule of activities to ensure mission success. Any real-time slips in schedule are quickly adjusted by the planners who, as ground controllers, are in constant communication with the crew.

Future astronauts on long duration exploration missions (LDEMs) will not have the same continuous, real-time communications with ground controllers that is available today. As such, astronauts will be expected to perform more autonomously—managing, prioritizing, and rescheduling their own schedules as they see fit (i.e., crew self-scheduling). However, astronauts are not experienced ISS planners, and these novice schedulers are expected to find self-scheduling a difficult task to complete due to the complexity of spaceflight scheduling. The challenge with crew self-scheduling is twofold: self-scheduling should not be burdensome to astronauts, and resulting timelines need to abide by the multiple requirements, resources, and constraints imposed by the mission. While it has been observed that self-scheduling is feasible in operational-like settings (Marquez, Hillenius, Healy & Silva-Martinez, 2019; Marquez et al., 2017), crew self-scheduling needs to be completed easily and quickly. However, "easily and quickly" is not necessarily compatible with scheduling task complexity that exists in human spaceflight missions.

Scheduling task complexity is driven by factors such as the number of activities to be scheduled with constraints. Spaceflight constraints range from simple to sophisticated (Dempsey, 2018). For example, some activities must be done exactly at a predetermined time, while other activities require particular resources (e.g., power and equipment) to be completed. Often, spaceflight activities have multiple constraints. The collection of activities and their associated constraints determines how hard it is to schedule and achieve an optimal timeline. There has been little research investigating the performance impacts of scheduling task complexity resulting from a lack of understanding of scheduling task performance and supportive mitigations to prevent nonoptimal performance. Our recent research (Lee et al., 2021) explored performance in scheduling and identified that activity constraints can impact task complexity and subsequently, associated performance. In the present study, human scheduling performance is investigated as a function of scheduling task complexity, specifically addressing the type and number of constraints as key factors.

Background: Scheduling, Complexity, and Human Performance

In the last several years, while research has been conducted that focuses on enabling increased flight crew autonomy for future LDEMs, there is comparatively little within the domain of autonomy in mission activity scheduling. Some research has emphasized solutions that support scheduling efficiency, such as decision support systems (e.g., Mishra et al, ., 2019) and algorithms for increased scheduling efficiency (e.g., Bu et al., 2016). Our previous research has centered around the design and usability of scheduling tools for self-scheduling in analog environments (Marquez et al., 2017, 2019) and in spaceflight (Marquez, Hillenius, Healy & Silva-Martinez, 2019). Others have explored the positive behavioral effects on crew with scheduling autonomy (Kanas, 2015; Roma et al., 2011). However, none of this research provides insight into the specific components of scheduling task complexity that exist in spaceflight operations and their effect on a scheduler's performance. Research is therefore required to address the current gap in the literature of complexity drivers of self-scheduling and associated mitigations and supportive mechanisms that enable crew members to self-schedule effectively and efficiently.

Self-scheduling research outside of the specialized spaceflight domain was also considered to further understand if self-scheduling complexity drivers and associated performance had been explored in other domains. Research on activity scheduling in wider domains, including surgery, nursing, construction, manufacturing, and power grids, provides some insight into the complexity drivers of scheduling and associated scheduling performance. There is a widely recognized desire to create an "optimal plan" in each domain, usually measured by domain-specific criteria of effectiveness (e.g., "how quickly patients are treated by nurses?") and efficiency (e.g., "are resources fully utilized?"; Erdogan & Denton, 2011). The complexity of the scheduling task is acknowledged within these varied domains and attributed to multiple levels of decision making and constraints between tasks (Argenziano et al., 2020). However, due to domain-specific complexity, research has dominantly focused on resolving the scheduling problem through various supportive mechanisms, including the development of modeling programs and automated scheduling tools (Chau et al., 2004). Some investigations in medical domains, including nursing and surgery, have focused on allowing self-scheduling as a solution, which has had a positive impact on wellbeing and satisfaction but provides no conclusive information on ensuring an optimal schedule as an outcome (Russell et al., 2012; Zhu et al., 2019). Additionally, self-scheduling nursing implementations have failed when team structures broke down and individuals prioritized their personal schedules over the success of the entire nursing staff (Bailyn et al., 2007).

Research on activity scheduling across domains has received an inconsistent focus on what individual factors drive complexity in the scheduling task and how they impact the creation of an optimal schedule. There is, therefore, a gap in understanding relating to complexity drivers for the planner, and the potential barriers to improving self-scheduling performance. Zhu et al. (2019) recently conducted a review of operating room planning and surgical case scheduling in which they discuss the literature around different types of constraints and problem features. One of the key types of constraints identified in the review was resource management, where a given resource (e.g., operating rooms) has a limited number, resulting in difficulties in scheduling and ultimately increasing patient waiting time and hospital-staff overtime. Commercial planning solvers were found to be insufficient to handle these types of constraints, requiring the development of genetic algorithms to find approximately optimal solutions in a timely manner for surgical scheduling (Erdem et al., 2012). Other types of constraints included the minimum/maximum time between two different events, such as requiring time between surgeries for disinfecting the operating rooms (Marques et al., 2012). Within the surgical domain, researchers have focused on different ways of minimizing the time between surgeries by scheduling patients with the same infection sequentially (Cardoen et al., 2009) or blocking infectious and noninfectious cases at different times (Hashemi Doulabi et al., 2016), though these techniques may not be broadly applicable to other domains. In addition to constraints, many planning techniques take the priority of the events into account. Castro and Marques (2015), for instance, divided their surgical activity scheduling into three priority levels based on the urgency and severity of the operations to be performed. Their scheduling algorithm prioritized the most urgent surgeries, scheduling them first, and worked down the priority list. Using real-life data as a case study, they were able to improve the occupation rate of their operating rooms and schedule more surgeries.

Overall, the lack of a consistent body of research into self-scheduling across domains has limited our understanding of how performance is affected by scheduling task complexity. The existing literature has investigated activity planning and scheduling with varied types and amounts of constraints, making it difficult to determine how each constraint influences performance. As a result, prior information from other domains has a limited contribution to understanding self-scheduling performance for long-duration spaceflight missions. In this research, we present a controlled, systematic evaluation of the individual types of constraints that are operationally relevant to spaceflight. To address some of the gaps in the literature, we examine human performance and resulting performance-influencing factors, including workload and situation awareness, of novice schedulers as a function of scheduling task complexity.

METHOD

Design

We investigated the effect of scheduling task complexity on scheduling performance by presenting participants with different types and numbers of constraints in a scheduling problem. We utilized a within-subject 4 × 2 experimental design which resulted in eight constrained scheduling problems and one unconstrained (baseline) scheduling problem for nine total problems. The two independent variables we manipulated were *type of constraint*, with four levels, and *number of constraints* with two levels, “low” (33% of activities to be scheduled with constraints) and “high” (66% of activities to be scheduled with constraints). A constraint was associated with an activity and defined a limitation or requirement that had to be met when the activity was scheduled. The *type of constraints* selected for this study are among those that currently exist in human spaceflight operations. Four types of constraints were presented:

- *Time Range Constraint (TR)* limits the time of day an activity can be scheduled (e.g., Activity A must start no earlier than 0900 and end no later than 1030);
- *Requires Constraint (R)* states that the activity needs to have a particular, static resource available (e.g., Activity requires communication availability);
- *Claim Constraint (CL)* describes a specific piece of equipment required for a particular activity or set of activities (e.g., Activities A and B both claim a treadmill, therefore, cannot be scheduled at the same time);
- *Ordering Constraint (O)* describes when an activity should be scheduled in relation to another activity (e.g., Activity A must be scheduled before Activity B).

Only one type of constraint was shown in any individual scheduling task (i.e., trials did not have multiple types of constraints). Scheduling problems either had a low number (33%) or a high number (66%) of constraints. The baseline problem was the first trial for all participants and contained no constraints. A Latin square determined the order for the remaining eight trials.

Participants

A power analysis was performed to determine the necessary sample size for a within-subject analysis with repeated measures and a very strong correlation among the repeated measures ($r = 0.80$; Evans, 1996). Using a medium effect size of $f = 0.25$ (Erdfelder et al., 1996), error probability (α) of 0.05, and power ($1 - \beta$) of 0.80 (Fritz & MacKinnon, 2007), the present framework required 16 subjects to detect within factors effects. Our previous research (Lee et al., 2021) indicated an effect size between medium and large, giving us confidence that 15 subjects would be sufficient for this study.

Fifteen individuals (seven females and eight males) volunteered for the experiment. While astronaut participants were not feasible, college educated participants were selected as a proxy for novice schedulers. Participants' ages ranged from 18–65. All held a minimum of a bachelor's degree, reported not to be color blind, and had previous experience using computer tablets. All were novice to the scheduling task and scheduling platform; "novice" was defined as having no current professional experience of scheduling, either as the main focus (e.g., mission scheduler) or as part of a current job role (e.g., project timeline manager). Demographics were collected on participants' experience with scheduling to ensure this criterion was met. Participants were recruited through advertisements requesting volunteers who met specific selection criteria. The study was approved by the NASA Ames Institutional Review Board (HRII 20-07).

Equipment and Materials

The experiment was administered remotely due to restrictions resulting from the COVID-19 pandemic. As such, participants were required to provide their own hardware: an iPad with access to Wi-Fi and a computer with microphone, speaker, and video camera. Participants accessed the scheduling platform, known as Playbook (Marquez et al., 2013; Marquez et al., 2017; Marquez, Hillenius, Zheng, et al., 2019) via iPad web browser; instructions and questionnaires were administered via computer web browser. The proctor ran a custom software platform that was developed to execute Playbook experimental trials and was used to collect the data (Kanefsky et al., 2018). Video conferencing was used on the computer for remote collaboration and to record the session.

Playbook

Playbook is a mobile, web-based scheduling software used to enable crew self-scheduling. In Playbook (Figure 1), the timeline view displays time horizontally from left to right with each crewmember having their own row and their own activities to be executed chronologically. An activity is displayed as a colored block with the length of the block directly related to the duration of the activity. Flexible activities (marked with a white dot) can be manipulated (i.e., scheduled and assigned by the user), and inflexible activities are fixed in time and cannot be moved. An activity may or may not have an associated constraint. If a constrained activity is scheduled and the constraint condition is not met, the activity is marked with a red outline denoting a constraint-based violation (Figure 1(d)). Overlapping activities are also flagged as a violation. For this experiment, all flexible activities had a scheduling priority (high, medium, or low priority). The task list view provides a list of the flexible activities and includes additional information such as priority level and associated constraints. The scratchpad facilitates the ability to move activities between the task list and the timeline and is located near the top of the Playbook interface (Figure 1(b)).

Scheduling Task

Participants were presented with a mostly empty timeline created for a one-day period. (For detailed description of activities, see Supplementary Material.) Similar to current astronaut schedules, part of the timeline already had inflexible scheduled activities, such as sleep and meals. Twenty-four flexible activities were positioned in the task list view at the beginning of each experimental trial; each activity had a priority (High, Medium, Low). Participants were required to select activities from the task list to move to the scratchpad, and then schedule them into the timeline (see Figure 1). There were no restrictions on the number of movements of each activity. More activities were listed in the task list than were possible to schedule to force participants to make a choice between priorities. Participants were trained to schedule by activity priority so that any activity left unscheduled should not be a higher priority than tasks that were scheduled. Participants were instructed to work as quickly as they could and schedule as many activities as possible without overlapping activities, which would create a violation. If a scheduled activity's constraint was not met, a constraint-based violation would also appear (see Figure 1(d)). Participants were also

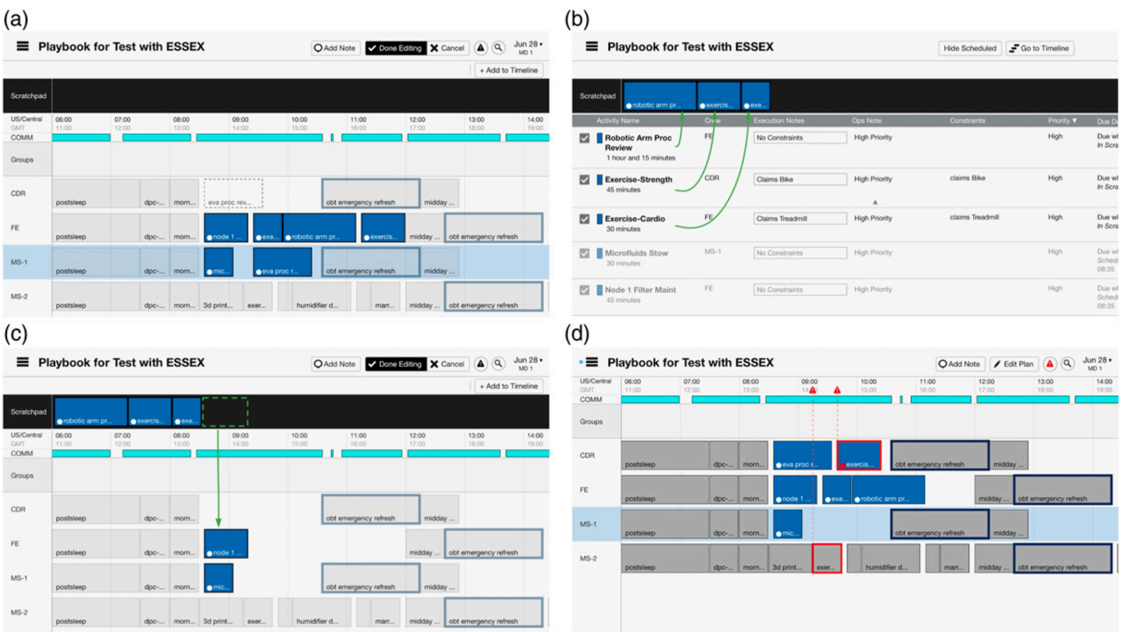

Figure 1. Playbook user interface, with gray inflexible activities, blue activities to be scheduled, and turquoise "COMM" availability depicted. Participants start with the initial plan (a) with only inflexible activities scheduled. From the task list (b), flexible activities can be moved to the scratchpad, where they are stored until the participant places the activities onto the timeline (c). If the participant creates a violation (d), the activities in violation are flagged with red outlines.

instructed to not to leave any violations in the timeline at the completion of the trial.

Measures

Dependent variables were selected based on findings from a pilot study (Lee et al., 2021). Task performance variables measured efficiency and effectiveness. The total time spent (referred to time on task) completing the scheduling task was recorded in minutes and used to infer efficiency. The number of constraint-based violations that were created by participants over the course of the scheduling task was used to measure effectiveness. Constraint-based violations were recorded regardless of whether the violation was subsequently resolved. As the baseline condition did not require activities with constraints to be scheduled, this condition was removed from the analysis of violations. In order to minimize individual differences and varied scheduling strategies (as seen in Lee et al., 2021), participants were asked to schedule quickly.

Situation awareness (SA) was also captured to provide insight into participants' knowledge of the environment and comprehension of requirements. Previous spaceflight research (Lee et al., 2021; Edwards et al., 2021) indicated that several critical components of effective scheduling are not relayed during formal training but instead come from experience. Experienced schedulers are aware of nonformal constraints, such as knowledge of physical space/layout or crew preference, and integrate this knowledge with their mental models when building schedules. Given that astronauts are not experienced ISS planners, understanding the level of SA a novice scheduler achieves can be used to identify barriers to establishing good SA and to inform the development of countermeasures to enhance SA for novice schedulers. At the end of every trial, participants were asked to respond to three SA-related questions, like the Situation Present Assessment Method (SPAM) technique (Durso et al., 1995), which had been modified for this study (Edwards et al., 2021). In line with the SPAM technique, participants were instructed to answer the SA questions as quickly and accurately as they could from what they remembered, though they could refer to the completed schedule, if desired. The "quickly and accurately" instruction was meant to address the speed-accuracy tradeoff. Someone who has a good awareness of the situation can answer questions quickly and easily from either memory or from within the display in comparison with someone who has not built that awareness. An accurate response alone is not necessarily meaningful; quickly and accurately indicates sufficient SA, while accurately with a long response time indicates the knowledge was not readily available.

After completing the SA questions, participants were asked to rate their workload on the NASA Task Load Index (NASA-TLX; Hart & Staveland, 1988). The traditional scale (1-100) was used. A pair-wise comparison questionnaire for the NASA-TLX subscales was also collected to identify the relevant weighting of the subscales to the overall NASA-TLX measure. This measure was selected to compare relative perceived workload per participant.

Protocol

Individuals who expressed interest in volunteering for the study were screened by phone to ensure they met the selection criteria. Informed consent forms were electronically signed, and demographic information was collected through an online form. On the day of the experiment, participants were given a standardized brief regarding the purpose of the study. Participants then took part in a training session, which included a video and four practice scheduling problems where they were asked to act as an astronaut tasked with developing schedules for oneself and one's crewmates. A nine-question competency check was conducted to verify participants were sufficiently trained. To take part in the experiment, participants were required to achieve a minimum of 77% (7/9 questions). A short break was provided between training and the experimental trials. After each trial, workload and SA questions were administered; a post experiment questionnaire and debrief was administered at the end of the experiment. Total participation time was approximately 2.5 hours.

RESULTS

Nonparametric analyses were used for statistical comparisons for most metrics. For our repeated-measures, Friedman's ANOVA was used; post-hoc comparison tests were conducted with Wilcoxon signed-rank tests. A series of four Wilcoxon tests were conducted to investigate the differences in violations between low (33%) and high (66%) number of constraints, for each type of constraint. Bonferroni correction was applied to create a corrected significance level of $p = 0.0125$.

When Friedman's analysis is reported, the main effect of *number of constraints* (holding type of constraint constant) is first reported, followed by the exploration of a main effect of *type of constraint* (holding number of constraints constant). For multiple comparisons, Bonferroni corrections were applied to set the appropriate significance level (e.g., four comparisons, significance level set at $p < 0.0125$). When a dependent variable met the assumptions for parametric data analysis (such as time on task and NASA-TLX score), repeated-measures ANOVA and pairwise comparisons for post-hoc tests were used.

Table 1 summarizes the descriptive statistics for the data analyzed. One dependent variable, time on task, met assumptions for parametric data analysis and we report mean (M) and standard deviation (*SD*) for each condition. All other variables did not meet parametric assumptions and we report median and interquartile range (IQR). Data from two trials (from different participants) were removed from the dataset due to technical issues collecting measures.

Performance-Influencing Factor: Self-Reported Workload

Friedman's ANOVA was utilized to enable a comparison analysis between the baseline and experimental conditions, followed by Wilcoxon signed-rank tests for post-hoc comparisons. Analysis of weighted NASA-TLX scores indicated no significant difference in self-reported workload between the baseline and the low *number of constraints* conditions. However, a significant difference was detected between the baseline and the high *number of constraints* conditions ($\chi^2(4) = 21.11$, $p < 0.001$, W = 0.35). Under this condition, workload was rated significantly lower in the baseline condition compared to most other constraint conditions: O ($Z = -2.96$, $p < 0.005$), CL ($Z = -3.18$, $p = 0.001$), and TR ($Z = -2.48$, $p < 0.05$), but not R ($Z = -0.22$, $p > 0.05$). See Table 1 for descriptive statistics.

An ANOVA analysis was used to evaluate the self-reported workload. A parametric analysis was utilized because NASA-TLX is considered a robust measure (Hart, 2006) and parametric analysis assumptions were statistically met. A significant main effect of the *number of constraints* was found on workload ($F(1, 12) = 6.16$, $p < 0.05$, $\eta_p^2 = 0.34$); on average, reported workload increased between low (M = 46.80) and high (M = 50.04) constraint conditions. In addition, a significant main effect of the *type of constraint* was identified ($F(3, 36) = 2.91$ $p < 0.05$, $\eta_p^2 = 0.2$). Post-hoc pairwise comparisons revealed no significant differences, although a finding that workload was higher in the CL condition (M = 51.01) than the R condition (M = 42.67) approached significance ($p = 0.09$). Finally, a significant interaction was found between *number of constraints* and *type of constraint* ($F(3, 36) = 3.06$, $p < 0.05$, $\eta_p^2 = 0.20$). This interaction causes the reported workload to be differentially affected by *type of constraint* in the high *number of constraints* condition compared to the low condition. Tukey post-hoc tests showed that reported workload was significantly higher for TR in the high condition compared to the low condition ($p = 0.0008$), while the other three constraints were unaffected. In the high constraints condition, pairwise comparisons also showed that R had significantly lower workload than O ($p = 0.0118$) and TR ($p = 0.0012$).

Performance-Influencing Factor: Situation Awareness

Response times to SA questions were not normally distributed, and thus, nonparametric analyses were used to identify statistical

Table 1. Summary of Descriptive Statistics for Dependent Measures Across Experimental Conditions. One Dependent Variable, Time on Task, Met Assumptions for Parametric Data Analysis, and we Report Mean and Standard Deviation (SD) for Each Condition. All Other Variables did not Meet Parametric Assumptions and Report Median and Interquartile Range (IQR).

	Dependent Variables				
Constraints	NASA-TLX	SA Percent Correct	SA Response Time (sec)	Time on Task (min)	Violations (Counts)
Baseline	49.67 (36.33)	66.67 (8.35)	10.67 (13.24)	4.52 (1.52)	N/A
Time range (TR)					
Low (33%)	42.67 (39.00)	66.67 (33.33)	8.54 (8.64)	6.08 (1.48)	6.00 (5.00)
High (66%)	59.00 (37.58)	100.00 (8.32)	9.72 (8.75)	7.37 (2.09)	10.00 (12.00)
Requires (R)					
Low (33%)	44.33 (32.33)	66.67 (66.67)	8.50 (8.81)	4.59 (1.42)	1.00 (4.00)
High (66%)	32.17 (22.67)	66.67 (33.33)	7.89 (3.98)	5.41 (1.30)	7.00 (4.25)
Claim (CL)					
Low (33%)	48.33 (43.67)	100.00 (66.67)	15.45 (11.15)	4.48 (1.17)	7.00 (6.00)
High (66%)	54.67 (39.33)	100.00 (33.33)	12.75 (30.19)	5.77 (1.33)	19.00 (18.00)
Ordering (O)					
Low (33%)	54.67 (27.67)	66.67 (33.33)	7.83 (33.83)	5.35 (1.62)	6.00 (8.00)
High (66%)	57.67 (36.67)	66.67 (33.33)	18.55 (26.62)	6.16 (1.83)	10.00 (8.00)

differences. Initial analyses explored all accurate responses, regardless of response time. Descriptive statistics for the percentage of correct responses and the average response times are provided in Table 1. Baseline data was excluded from this analysis.

No statistical difference was detected with respect to SA percentage of correct responses as a function of *type of constraint* or *number of constraints*. Overall, 75% of SA questions were answered correctly. There was, however, a statistically significant difference in SA response times within the high *number of constraints* condition ($\chi^2(3) = 12.50, p < 0.01$, W = 0.2), but not for low *number of constraints* condition ($\chi^2(3) = 6.28, p = 0.08$). For the high *number of constraints* condition, accurate SA response times were longer for O compared to TR ($Z = -3.04, p < 0.01$) or R ($Z = -3.04, p < 0.01$) and longer for CL compared to R ($Z = -1.98, p < 0.05$).

Similar to the SPAM technique, response times greater than 40.5 seconds were re-coded as incorrect. The assumption is that participants had not gained sufficient SA to answer the question, but instead were relying on access to the completed schedule to answer questions. As a result, 20 individual data points (6% of the dataset) were removed and analyses re-run. With this revised dataset, no statistical difference was detected with respect to SA percentage of correct response as a function of *number of constraints.* A statistically significant difference in percent correct was detected for the high *number of constraints* conditions ($\chi^2(3) = 15.63, p < 0.001$, W = 0.4), and approached significance for low *number of constraints* ($\chi^2(3) = 6.85, p = 0.08$). For the high *number of constraints* conditions, the percent correct was lowest for O compared to TR ($Z = -2.88, p < 0.01$) or R ($Z = -2.92, p < 0.01$).

No statistical difference was detected in the response time necessary to accurately answer SA questions for either of the independent variables. While initial analyses detected a difference in response times for high *number of constraints*, this effect disappeared once a time limit was placed on accurate responses. Seventy percent (14/20) of the individual data points that exceeded the time limit came from the O constraint.

Efficiency: Time on Task

Time on task met assumptions for parametric data analysis, and hence, ANOVA results and Bonferroni pairwise comparisons are reported. Descriptive statistics (Table 1) suggest that the baseline condition (where no activities contained constraints) was completed faster than all other experimental conditions except R-low and CL-low, which were on par. To detect significant differences in time on task between the baseline condition and experimental conditions, eight paired samples t-tests were conducted. Using a Bonferroni correction ($p < 0.00625$), only two comparisons were significant. On average, participants took less time to complete the baseline condition in comparison to O-high ($p < 0.001$) and TR-high ($p = 0.001$) conditions.

For the experimental conditions (Figure 2), a significant main effect of the *number of constraints* was found on time on task ($F(1, 12) = 28.86$, $p < 0.001$, $\eta_p^2 = 0.7$), with the high condition (M = 6.19 minutes) on average taking longer to complete than the low condition (M = 4.98 minutes). In addition, a significant main effect of the *type of constraint* was identified ($F(3, 36) = 6.17, p < 0.005, \eta_p^2 = 0.34$). Post-hoc Bonferroni pairwise comparisons revealed that participants took significantly longer to complete TR trials than R trials ($p < 0.01$). TR trials were also longer than CL trials, which approached significance ($p = 0.06$). No significant interaction effects were identified.

Effectiveness: Violations of Constraint-Based Rules

Using Wilcoxon signed ranks tests and adjusting the significance level, the effect of *number of constraints* were analyzed for each constraint type. Descriptive statistics (Table 1) suggest that there were fewer violations created when there were fewer constraints. For each *type of constraint*, there was a significant difference in the number of violations between the low and high *number of constraints*: O ($Z = -2.84, p = 0.005$), CL ($Z = -3.36, p = 0.001$), TR ($Z = -2.61, p < 0.01$) and R ($Z = -3.12, p < 0.005$). Results indicate a significant main effect of the *number of constraints* on activities to be

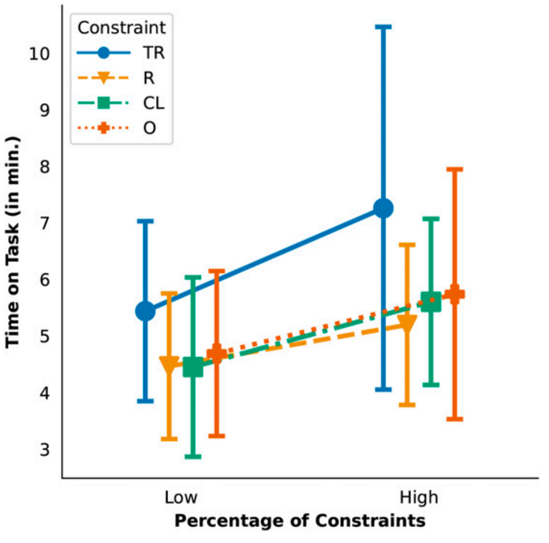

Figure 2. Time to complete the scheduling task (time on task), in minutes, by type of constraint and number of constraints. Data points are the means, and the error bars are the standard deviation.

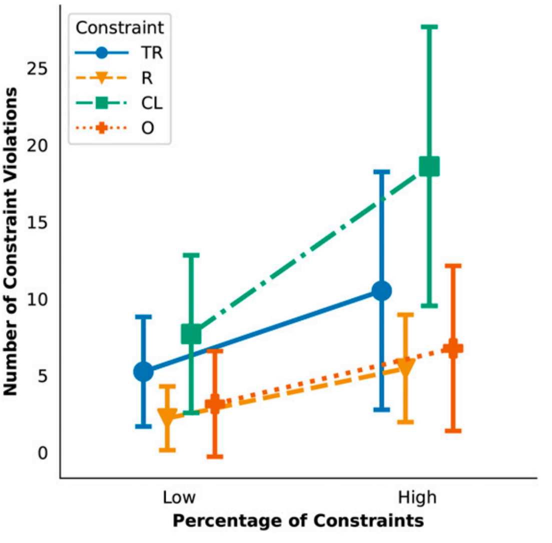

Figure 3. Number of constraint violations created during the scheduling task by type of constraint and number of constraints. Data points are the means, and the error bars are the standard deviation.

scheduled on the number of constraint-related violations, with more violations created for the high condition (Figure 3). When assessing the effect of *type of constraint*, the number of violations was significantly different for both the low number of constraints condition ($\chi^2(3)$ = 14.85, $p < 0.005$, W = 0.33) and the high ($\chi^2(3)$ = 11.93, $p < 0.01$, W = 0.31). For the low condition, the number of violations created in R trials was significantly less than violations in CL trials (Z = −3.11, $p < 0.005$) and TR trials (Z = −3.14. $p < 0.005$). Fewer violations in R compared to O trials approached significance (Z = −2.03, p = 0.04) (in accordance with the Bonferroni corrected significance level of p = 0.0125). In the high condition, violations in the CL condition were significantly higher than R (Z = −2.90, $p < 0.005$) and O trials (Z = −2.39, $p < 0.05$).

DISCUSSION

A within-subjects design was used to investigate the effect of *number of constraints* and *type of constraint* on human performance when creating a schedule. We examined workload, situation awareness, efficiency and effectiveness of novice schedulers performing this task. Performance efficiency was measured by how long it took participants to create a schedule, and performance effectiveness was inferred from the number of violations participants created in the process of scheduling activities. This study contributes to the current research gap of understanding complexity drivers in activity scheduling and associated scheduling performance.

Reported workload was affected by *type of constraint* as well as *number of constraints* to be scheduled. Compared to baseline, participants reported workload to be significantly greater in the high *number of constraint* condition but not significantly greater in the low *number of constraint* condition, suggesting that the number of constraints affected workload regardless of *type of constraint*. The lack of significant difference in reported workload between baseline and the low constraint conditions also suggests that participants found this condition manageable, and the increased task complexity was not influential on perceived workload. A main effect of *type of constraint* on reported workload was also found, although subsequent comparisons did not find significant differences between the four constraint types, suggesting a smaller effect on reported workload compared to *number of constraints*. Nonetheless, more constraints resulted in higher reported workload. This is consistent with cognitive resource theories, where more resources are required for harder tasks (Smit et al., 2004; Ackerman et al., 1984; Bellenkes et al., 1997). The low constraint condition may have been tolerated by participants with spare capacity remaining, explaining the lack of significant difference with baseline. However, the high number of constraint condition demanded more resources of participants compared to either the baseline or low constraint condition, resulting in an increased perception of workload.

The adaptation of the SPAM methodology was successful in measuring differences in SA in participants across conditions. Participants were asked questions that probed their perception and prediction of elements within the timeline they had scheduled. Similar to Edwards et al. (2021) an effect for the *type of constraint* was observed for SA response time. Surprisingly, no difference was detected between low and high *number of constraints*, though the effect of constraint type was larger in the latter condition. This suggests that SA is more greatly affected by the *type of constraint* as opposed to the *number of constraints*. Limitations of SPAM have been documented in the literature which may have influenced results, such as speed-accuracy tradeoffs have been documented in the literature, and a potential for results to be biased due to more accurate and faster answering during low workload periods (Endsley, 2021). However, some of these documented limitations were addressed in the present study due to the nature of the modified version of SPAM that was used (see Method section) (e.g., Durso et al., (1995)).

Regarding time on task and number of violations, there were main effects due to the *number of constraints* and the *type of constraint*. More constraints resulted in more time spent scheduling and more violations created. While this result is consistent with Lee et al. (2021) this experiment shows that this effect is seen regardless of the type of constraint. When participants had to schedule more activities with constraints, the scheduling task became harder to complete. Having more constraints to meet resulted in a higher number of total constraint-based violations created while scheduling. This implies that novice schedulers sometimes attempt to schedule activities at the times where the activity is constrained (i.e., creating invalid schedule).

We provided cues to the participants that would enable them to identify times where activity could be scheduled to meet a constraint. All the constraints were described in the instructions (e.g., “Requires COMM”), which were always visible on a second browser during the trial; this same information was provided in the Task List; and finally, if the activity was selected in Playbook, the constraint description was visible. Participants may have chosen a trial-and-error approach in trying to schedule activities with constraints, leading to more violations as well. In turn, resolving violations elongated the amount of time spent creating a schedule that satisfies the given constraints.

Interestingly, when compared to scheduling activities with no constraints (i.e., baseline), performance was comparable to that when scheduling in the low *number of constraints* conditions. There was no significant difference in how long it took participants to complete the baseline trial compared to the low constraint conditions. Similarly, no significant difference in perceived workload was detected between the baseline and low number of constraints conditions. Even if participants chose a trial-and-error approach, requiring an additional physical act of moving around activities with constraints and mental challenge of identifying an adequate violation-free schedule, no significant differences in reported workload were identified. Considering these two findings together, our results suggests that a few constraints does not significantly impact scheduling performance, nor does it add significant workload.

These results suggest that a scheduling task with 66% of activities having constraints falls under the hard problem category, but not the over-constrained problem category (Cheeseman et al., 1991). As Cheeseman et al. (1991) suggest, under- and over-constrained problems are easy because there are correspondingly many or very few solutions. Hard problems are those in some critical region of difficulty (which is problem dependent). Similarly, participants spent a longer amount of time, effort, and incurred more trial-and-error attempting to find a valid schedule, of which there are a limited number.

Decreased performance and increased workload was detected when participants were asked to schedule more activities with constraints (i.e., the high *number of constraints* conditions). When there are more constraints to meet, the problem space (or valid schedule solutions) becomes smaller, requiring participants to spend more time and effort finding a valid scheduling solution. Essentially, the possible locations (or start times) in which activities can be scheduled is limited, and participants spend more time rescheduling multiple activities to find a valid schedule. In turn, finding a valid schedule solution in the low *number of constraints* conditions is not as challenging to participants.

Differences in Types of Constraints

While there is an effect due to type of constraint, there is mixed evidence to indicate which constraint was hardest for participants to schedule. The Time Range constraint was the hardest for participants to schedule as those trials took the longest to complete regardless of number of constraints. Most prominently, there was a significant interaction between *number of constraints* and *type of constraint* affecting reported workload. Post-hoc comparisons indicate that the main contributor was the condition where participants had to schedule activities where 66% of them had an associated Time Range constraint (TR-high). For this condition, participants reported the highest workload. This suggests that the increase in constraints affected the Time Range task complexity greater than other constraints. The Time Range constraint required activities to either be scheduled during a morning or an afternoon two-hour time slot for any of the three crew member timelines. Participants might have found this constraint the most restrictive (i.e., limited flexibility existed to schedule these activities with this constraint).

When considering performance effectiveness, the constraint that led to the most violations was Claim, particularly for the high *number of constraints* conditions. The median number of violations for 66% Claim (CL-high) was almost twice that of the other three constraint types (Table 1). However, this did not result in changes in performance regarding efficiency or workload (though less SA was developed in comparison to Requires constraints). Solving for the Claim constraint required participants to understand which other activities claimed the same resource. This required participants to consider the scheduling of at most four activities at once so that they would not be scheduled concurrently. While this led to more violations, it did not increase participant's reported workload, nor did it result in longer time to schedule all activities. Participants may have found it more challenging to identify ways to meet the constraint but there was sufficient scheduling flexibility to accommodate the activities in the timeline.

The Ordering constraint required participants to understand how at least two activities were scheduled. To satisfy this constraint, the activities must be scheduled in the appropriate order (e.g., setup activity is before cleanup). Situation awareness response time for Ordering was significantly longer compared to Time Range and Requires. When a response time cutoff was applied to the dataset, participants' SA percentage of correct response was again significantly poorer than Time Range and Requires. Additionally, the majority of the response times that were longer than 40 seconds were during the Ordering trials. These results suggest that SA was either the lowest or most undeveloped for Ordering constraint. Interestingly, scheduling performance (as measured by number of violations and time on task) was not significantly affected by poor SA. Overall, while the task of scheduling activities with this type of constraint was feasible, it appears that conceptualizing a mental model of a constraint network may be challenging for novice schedulers. This is particularly noteworthy as only a simple constraint network (just two activities) was used in this experiment and more complex networks (three or more activities) are commonly found in spaceflight operations.

An interesting trend was identified for only the Requires constraint condition. For activities with this constraint, participants were asked to schedule the activity at a time where the required resource was available, namely, when there was sufficient communication availability. While scheduling performance increased between low and high number of constraints for the Requires constraint, workload and SA response time tended to decrease. One possible explanation for this trend (and significant interaction for workload) is that the direct presentation of communication availability in the user interface (as seen in Figure 1) allowed participants to more easily identify where activities could be scheduled to satisfy the constraint. This would also explain why performance for Requires trials are consistently one of the better ones (quicker to schedule and lower number of violations). The visual presentation of communication availability might have been sufficient to decrease perceived workload in the high constraint condition despite having to do a more challenging scheduling task.

The mixed results across the type of constraints studied may be an indicator that activity scheduling support strategies differ across the type of constraints. Descriptions of the constraints are not sufficient for novice schedulers. The trends observed for Requires suggest that visual aids may be most useful. Time Range and Claim constraints might benefit the most from a visual scheduling aid. Finally, Ordering constraints may be too challenging for novice schedulers to develop and internalize a mental model sufficient to develop situational awareness.

Summary

Overall, this experiment suggests that *number of constraints* and *type of constraints* are key factors in scheduling task complexity and human performance for novice schedulers. Efficiency, effectiveness, and reported workload increase with the number of constraints, though schedules with a lower number of constraints do not significantly impact performance in comparison to schedules with no constraints. Differences in situation awareness were detected for *type of constraint*, though this appears to be mainly driven by one specific type of constraint. While there is a main effect due to *type of constraints* based on performance, there is evidence to suggest that the various constraints affected performance in different manners. The results imply that novice schedulers adequately perform scheduling with lower number of constraints and would benefit from scheduling aids to minimize the decreased performance observed with higher number of constraints.

The findings from this study contribute to the current gap in activity scheduling performance literature by systematically investigating self-scheduling in novice schedulers in a controlled, laboratory environment. Results provide specific insight into the components of self-scheduling that can affect workload and scheduling performance in crew and may therefore inform the creation of future mitigation strategies to specifically address the main drivers in self-scheduling complexity. Thus, this work provides an important initial step in enabling future LDEMs.

Future Directions

Future work will focus on investigating if there are objective measures to detect individual and strategy differences across participants. While participants were all given the same training, it is likely that a few distinct scheduling strategies emerged. They might provide insight as to which specific activity constraints were more challenging to schedule. In turn, this will inform the development of future scheduling aids, be it visualizations or planning practices, for the most challenging constraint types.

Practical Applications

The research results presented contribute to our understanding of novice schedulers' performance. They help identify areas of scheduling task complexity, which in turn leads to informing the type and design of decision aids for schedules. Results will enable future self-scheduling for LDEM. Furthermore, results contribute to the body of knowledge for other complex scheduling conducted in nursing, surgery, and manufacturing domains. Findings suggest the importance of novices having sufficient time and training to develop adequate situation awareness regarding complex constraints.

KEY POINTS

- Changes in human performance was measured as a function of scheduling task complexity for novice schedulers.
- A main effect was detected for the number of constraints and type of constraint for efficiency, effectiveness, and workload. A significant interaction was detected for workload.
- Novice schedulers adequately perform scheduling with a lower number of constraints at a level comparable to an unconstrained scheduling problem.
- The type of constraints affected performance in different ways, suggesting targeted scheduling aids based on constraint type could be effective.

ACKNOWLEDGMENTS

The authors would like to thank Nick Bathurst for his comments in developing this manuscript. Correspondence concerning this article should be addressed to Jessica J. Marquez, Ph.D. (jessica.j.marquez@nasa.gov).

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: 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.

SUPPLEMENTARY MATERIAL

Supplemental material for this article is available online.

REFERENCES

Ackerman, P. L., Schneider, W., & Wickens, C. D. (1984). Deciding the existence of a time-sharing ability: A combined methodological and theoretical approach. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(1), 71–82. https://doi.org/10.1177/001872088402600107

Argenziano, M., Fischkoff, K., & Smith, C. R. (2020). Surgery scheduling in a crisis. *New England Journal of Medicine*, *382*(23), e87. https://doi.org/10.1056/NEJMc2017424

Bailyn, L., Collins, R., & Song, Y. (2007). Self-scheduling for hospital nurses: An attempt and its difficulties. *Journal of Nursing Management*, *15*(1), 72–77. https://doi.org/10.1111/j.1365-2934.2006.00633.x

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, *68*(7), 569-579.

Bu, H.-J., Zhang, J., & Luo, Y.-Z. (2016). Constraint satisfaction and optimization for space station short-term mission planning based on an iterative conflict-repair method. *Engineering Optimization*, *48*(10), 1658–1678. https://doi.org/10.1080/0305215X.2015.1137566

Cardoen, B., Demeulemeester, E., & Beliën, J. (2009). Sequencing surgical cases in a day-care environment: An exact branch-and-price approach. *Computers & Operations Research*, *36*(9), 2660–2669. https://doi.org/10.1016/j.cor.2008.11.012

Castro, P. M., & Marques, I. (2015). Operating room scheduling with generalized disjunctive programming. *Computers & Operations Research*, *64*, 262–273. https://doi.org/10.1016/j.cor.2015.06.002

Chau, K. W., Anson, M., & Zhang, J. P. (2004). Four-dimensional visualization of construction scheduling and site utilization. *Journal of Construction Engineering and Management*, *130*(4), 598–606. https://doi.org/10.1061/(ASCE)0733-9364(2004)130:4(598)

Cheeseman, P., Kanefsky, B., & Taylor, W. M. (1991). Where the really hard problems are. Proceedings of the 12th International Joint Conference on Artificial Intelligence -Volume 1, 331–337.

Dempsey, R. C. (Ed.) (2018). *The International Space Station: Operating an Outpost in the New Frontier*. National Aeronautics and Space Administration.

Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M., Nikolic, D., Moertl, P. M., Ohrt, D., & Manning, C. A. (1995). Expertise and chess: A pilot study comparing situation awareness methodologies. In D. Garland & M. R. Endsley (Eds.), *Proceedings of the International Conference on Experimental Analysis and Measurement of Situation Awareness* (pp. 295–303).

Edwards, T., Brandt, S., & Marquez, J. (2021). Towards a measure of situation awareness for space mission schedulers. In AHFE 2021 International Conference. Virtual Event. Springer.

Endsley, M. R. (2021). A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *63*(1), 124–150. https://doi.org/10.1177/0018720819875376

Erdem, E., Qu, X., & Shi, J. (2012). Rescheduling of elective patients upon the arrival of emergency patients. *Decision Support Systems*, *54*(1), 551–563. https://doi.org/10.1016/j.dss.2012.08.002

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods,* *Instruments, & Computers*, *28*(1), 1–11. https://doi.org/10.3758/BF03203630

Erdogan, S. A., & Denton, B. T. (2011). Surgery planning and scheduling. In J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, & J. C. Smith (Eds.), *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470400531.eorms0861

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Hashemi Doulabi, S. H., Rousseau, L.-M., & Pesant, G. (2016). A constraint-programming-based branch-and-price-and-cut approach for operating room planning and scheduling. *INFORMS Journal on Computing*, *28*(3), 432–448. https://doi.org/10.1287/ijoc.2015.0686

Kanas, N. (2015). Autonomy and the crew-ground interaction. In N. Kanas (Ed.), *Humans in Space* (pp. 99–107). Springer International Publishing. https://doi.org/10.1007/978-3-319-18869-0_9

Kanefsky, B., Zheng, J., Deliz, I., Marquez, J. J., & Hillenius, S. (2018). Playbook data analysis tool: Collecting interaction data from extremely remote users. In T. Ahram, & C. Falcão (Eds.), *Advances in Usability and User Experience* (pp. 303–313). Springer International Publishing. https://doi.org/10.1007/978-3-319-60492-3_29.

Lee, C., Marquez, J., & Edwards, T. (2021, January 11). Crew autonomy through self-scheduling: Scheduling performance pilot study [Paper presentation]. AIAA Scitech 2021 Forum, VIRTUAL EVENT. https://doi.org/10.2514/6.2021-1578

Marques, I., Captivo, M. E., & Pato, M. V. (2012). Planning elective surgeries in a Portuguese hospital: Study of different mutation rules for a genetic heuristic. *Lecture Notes in Management Science*, *4*, 238-243.

Marquez, J. J., Hillenius, S., Healy, M., & Silva-Martinez, J. (2019). *Lessons Learned from International space station crew autonomous scheduling test*: International Workshop of Planning and Scheduling for Space (IWPSS).

Marquez, J. J., Hillenius, S., Kanefsky, B., Zheng, J., Deliz, I., & Reagan, M. (2017). Increasing crew autonomy for long duration exploration missions: Self-scheduling [Paper presentation]. 2017 IEEE Aerospace Conference (pp. 1–10), Big Sky, MT, IEEE. https://doi.org/10.1109/AERO.2017.7943838

Marquez, J. J., Hillenius, S., Zheng, J., Deliz, I., Kanefsky, B., & Gale, J. (2019b). Designing for astronaut-centric planning and scheduling aids. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 468–469. https://doi.org/10.1177/1071181319631386

Marquez, J. J., Pyrzak, G., Hashemi, S., McMillin, K., & Medwid, J. (2013). Supporting real-time operations and execution through timeline and scheduling aids [Paper presentation]. 43rd International conference on environmental systems (Vol. 1). American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2013-3519

Mishra, A. P., Sengupta, S., Sreedharan, S., Chakraborti, T., & Kambhampati, S. (2019). *Cap: A decision support system for crew scheduling using automated planning*. Naturalistic Decision Making.

Roma, P. G., Hursh, S. R., Hienz, R. D., Emurian, H. H., Gasior, E. D., Brinson, Z. S., & Brady, J. V. (2011). Behavioral and biological effects of autonomous versus scheduled mission management in simulated space-dwelling groups. *Acta Astronautica*, *68*(9–10), 1581–1588. https://doi.org/10.1016/j.actaastro.2009.09.034

Russell, E., Hawkins, J., & Arnold, K. A. (2012). Guidelines for successful self-scheduling on nursing units. *JONA: The Journal of Nursing Administration*, *42*(9), 408–409. https://doi.org/10.1097/NNA.0b013e3182664dd8

Smit, A. S., Eling, P. A. T. M., & Coenen, A. M. L. (2004). Mental effort affects vigilance enduringly: After-effects in EEG and behavior. *International Journal of Psychophysiology*, *53*(3), 239–243. https://doi.org/10.1016/j.ijpsycho.2004.04.005

Zhu, S., Fan, W., Yang, S., Pei, J., & Pardalos, P. M. (2019). Operating room planning and surgical case scheduling: A review of literature. *Journal of Combinatorial Optimization*, *37*(3), 757–805. https://doi.org/10.1007/s10878-018-0322-6

Jessica J. Marquez is a human-systems engineer at NASA Ames Research Center. She holds a PhD (2007) and a master's from MIT (2001), and specializes in space human factors engineering. She currently leads the Scheduling and Planning Interface for Exploration (SPIFe) team within the Human-Computer Interaction Group.

Tamsyn Edwards is a Senior Research Associate with San Jose State University at NASA Ames. She holds a PhD in Human Factors from the University of Nottingham, UK, in collaboration with EUROCONTROL (2013).

John A. Karasinski is a Research Scientist in the Human Systems Integration Division at NASA Ames Research Center. He holds a bachelor's in Physics from the University of California, Santa Cruz (2012), and a master's and PhD in Mechanical and Aerospace Engineering from the University of California, Davis (2016, and 2020).

Candice N. Lee is a research intern with the Human-Computer Interaction Group at NASA Ames Research Center. Currently, she is a Human Factors Engineer intern at Genentech. She holds a bachelor's in Psychology (2018) and a master's in Human Factors from San Jose State University (2021).

Megan C. Shyr is a Student Trainee in the Human Systems Integration Division at NASA Ames Research Center. She has a master's in Mechanical and Aerospace Engineering (MAE) from the University of California, Davis (2018) and is currently pursuing her PhD (MAE, UC Davis).

Casey L. Miller is a research intern with the Human-Computer Interaction Group at NASA Ames Research Center. She has a bachelor's in Aerospace Science and Engineering from the University of California, Davis (2017) and is currently pursuing her master's in Mechanical and Aerospace Engineering (UC Davis, 2022).

Summer L. Brandt is a Research Psychologist in the Human Systems Integration Division at NASA Ames Research Center. She holds a master's degree in Experimental Psychology from San Jose State University (2003).

Date received: June 11, 2021
Date accepted: October 11, 2021