# ENABLING ASTRONAUT SELF-SCHEDULING USING A ROBUST MODELLING AND SCHEDULING SYSTEM (RAMS): A MARS ANALOG USE CASE

Michael Saint-Guillain[1], Jean Vanderdonckt[1], Nicolas Burny[1], Vladimir Pletser[2], Tiago Vaquero[3], Steve Chien[3], Alexander Karl[4], Audrey Comein[1†], Cheyenne Chamart[1†], Cyril Wain[1†], Ignacio S. Casla[1†], Jean Jacobs[1†], Julie Manon[1†], Julien Meert[1†], Sirga Drouet[1†]

[1] Université catholique de Louvain (UCLouvain, Belgium)

[2] European Space Agency, ret.; Blue Abyss (UK)

[3] Jet Propulsion Laboratory, California Institute of Technology (Caltech, CA, USA)

[4] Space Applications Services (Belgium)

## ABSTRACT

Human long duration exploration missions (LDEMs) raise a number of technological challenges. This paper addresses the question of the crew autonomy: as the distances increase, the communication delays and constraints tend to prevent the astronauts from being monitored and supported by a real time ground control. Eventually, future planetary missions will necessarily require a form of *astronaut self-scheduling*. We study the usage of a computer decision-support tool by a crew of analog astronauts, during a Mars simulation mission conducted at the Mars Desert Research Station (MDRS, Mars Society) in Utah. The proposed tool, called *Romie* [1], belongs to the new category of *robust modelling and scheduling* (RAMS) systems. It allows the crew members (i) to visually model their scientific objectives and constraints, (ii) to compute near-optimal operational schedules while taking uncertainty into account, (iii) to monitor the execution of past and current activities, and (iv) to modify scientific objectives/constraints w.r.t. unforeseen events and opportunistic science. In this study, we empirically measure how the astronauts, which are novice planners, perform at using such tool when self-scheduling under the realistic assumptions of a simulated Martian planetary habitat.

## 1. INTRODUCTION

Past space missions have had very limited experience in human self-scheduling. In fact, [2] states that current human operations, including extravehicular activities (EVAs), are *"carefully choreographed, and rehearsed events, planned to the minute by a large team of EVA engineers, and guided continuously from Earth"* [3, 4]. As the distances increase however, the communication delays rapidly become an obstacle to remote real time monitoring and management of operations from Earth. However, human operations on Mars are expected to be carried at a faster rate than current rover missions[1] [5], which implies new planning strategies and tools that account for latency-impacted interactions [6]. In addition, future planetary EVAs are likely to be driven by science [7], requiring flexible adaptations according to scientific samples. In such context, future human space missions will have to enable some degree of crew autonomy and self-scheduling capabilities.

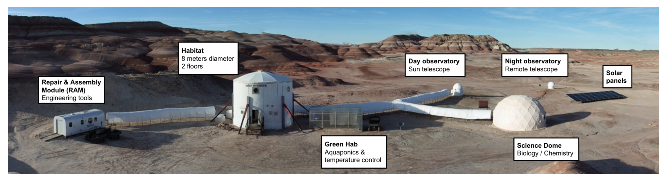


**Fig. 1**. The Mars Desert Research Station (MDRS), located in the Utah desert, is a Mars analog planetary habitat (Mars Society).

The problem of scheduling a set of operations in a constrained context such as the *Mars Desert Research Station* (MDRS, Fig. 1) is not trivial, even in its classical deterministic version. It is a generalization of the well-known *job-shop scheduling problem*, which has the reputation of being one of the most computationally demanding problems [8]. In [9], the authors raise on the importance of mission planning, as 25% of the budget of a space mission may be spent in making these decisions beforehand, citing the Voyager 2 space probe for which the development of the a priori schedule involving around 175 experiments requiring 30 people during six months. Nowadays, hardware and techniques have evolved and it is likely that a couple of super-equipped (*i.e.* with a brand new laptop) human brains may suffice in that specific case. Yet, the problems and requirements have evolved too. Instead of the single machine Voyager 2, space missions have to deal with teams of astronauts.

† These authors contributed equally.

[1]Current Mars rover missions are commanded by the ground operations team at most once per Martian day, or sol, and operate independently in between such contacts.

### 1.1. Rescheduling on-the-fly: objectives, constraints and opportunistic science

A human mission on Mars is different will necessarily be a long duration exploration mission (LDEM). The communication delays, in each direction, range from 3 to 22 minutes. Finally, in the current configuration of Mars orbiters, only a few short communication windows with Earth are possible per each Martian day (called a *sol*), with limited data rate (2 Mega bits per second).

In such conditions, any deviation from the original plan must be managed on the fly by the astronauts themselves. However, [10] demonstrated the fact that astronauts are not good at solving such complex problems by hand. This is not surprising. The sheer complexity of space systems means that thousands of constraints must be accounted for in decision making, and balancing of a large number of competing soft objectives must also be considered. An articulation of the size of this problem space for the Rosetta Orbiter mission science planning is described in [11] and a future human mission to Mars is likely to be orders of magnitude more complex. Furthermore, the astronauts must also be able to adapt their schedules according to new scientific goals and requirements, such as conducting opportunistic science (*e.g.*, recording a dust devil), or even a new scientific project, or unexpected events such as machine breakdowns. In other words, the human machine team must be able to track evolving scientific objectives and operations constraints to re-optimize activities in an ever changing mission context.

### 1.2. Robust Advanced Modelling and Scheduling (RAMS)

A recent review of planning and scheduling tools, specific of applied to either space exploration and industrial operations is provided in [1]. Existing systems usually fall into *a)* being specifically designed for a particular application/mission or operational context, or *b)* not having a generic, integrated optimization system to generate robust schedules (from a probabilistic point of view). Instead, the *Romie* RAMS system is used in this study. Compared to classical frameworks, a RAMS system such as provides the following *technological innovations*:

1. Graphical problem modelling. The user is able to graphically draw and manipulate the structure and constraints of its scheduling problem, including stochastic models for task durations.

2. Optimization under uncertainty. An optimization engine allows the user to generate, or adapt existing schedules, in a way that produces schedules robust w.r.t. uncertainty.

Unlike all existing tools, in a (robust) advanced modelling and scheduling (RAMS) system, both modelling and modifying the problem is now made accessible to the end-user, which is critical for a reliable self-scheduling. Although being a hot research domain, *Romie* is the first scheduling tool to propose an integrated robust (*i.e.* under uncertainty) optimization engine. Having more robust (*i.e.* reliable) schedules, the end users are more likely to avoid last minute rescheduling. *Eventually, what-if analysis, as well as sensitivity analysis, become less relevant*: the solutions are optimized following directly the KPIs *expected values* and considering the uncertainties related to task execution.

We believe that both points *1.* and *2.* provide significantly more autonomy to users, whom remain otherwise highly dependent of planning and scheduling experts. Based on the theoretical foundations defined in [12], the empirical contribution of point *2.* has been extensively validated in [13, 1, 14]. Testing the ability of the non-experts end-users to actually "self-schedule" using *1.* is the main goal of this study.

## 2. THE M.A.R.S. UCLOUVAIN 2022 MISSION

Our study on astronaut self-scheduling is driven by the scientific research projects to be carried out by the crew members in the context of the simulation. Before the actual beginning of the mission, the selected projects have been modelled in the *Romie* system, and provisional schedules have been designed.

### 2.1. Experimental plan

Several scientific research projects are to be conducted at the MDRS. Each project will be carried on in place by either one or two astronauts, and some projects (such as health projects) involve the participation of all the crew members. Yet, these projects are designed and prepared months ahead. During that period, preliminary experiments can be conducted on *Romie*, in order to get first results on the system's usage by the astronaut, in offline (supervised) conditions. The actual *M.A.R.S. UCLouvain 2022* mission period, which lasts 12 days on field at the MDRS (see Figure 1), constitutes the main material of this study. Day after day, each crew member will use the *Romie* system to monitor and update their operations. Measures using different techniques will be made each time the astronauts use the system, in order to record and further analyze how well they succeed at self-scheduling.

What is the on-boarding time of the RAMS? Are the decisions taken by the crew members of acceptable quality, and how long does it take for a novice user before setting up correct schedules? Are our astronauts all able to adapt their scientific objectives and schedules as the operations evolve? Our study tackles these questions by focusing on the temporal evolution of these following two complementary KPIs: system usability and user experience. The ISO-9241-210 standard [15] defines the usability as *the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*. User Experience is defined by the same standard as *the user's perceptions and responses that result from the use and/or anticipated use of a system, product or service* and is generally understood as inherently dynamic, given the ever-changing internal and emotional state of a person and differences in the circumstances during and after an interaction with a product [16].

### 2.2. The Mars Desert Research Station

The MDRS in the desert of Utah has been in operation since 2002 from November through April every year. The geologic features of the surrounding Jurassic–Cretaceous terrain also make the desert environment seem Mars-like to crew members. The MDRS habitat itself is a vertical cylindrical structure of approximately 6 m diameter and 8 m high, composed of two floors. The ground floor (lower deck) includes a front door airlock used for simulated EVA, an EVA preparation room,a large room used as a laboratory for geology and biology activities, a small engineering workshop area, a second back door airlock for engineering activities, a small bathroom and a toilet, three small windows, and a stair leading to the first floor. The first floor (upper deck) includes a common area or living room with a central table, a wall-attached circular computer/electronic table, a kitchen corner, six small bedrooms, and a loft on top of the small bedrooms. Some panoramic pictures from the inside at provided in Figure 2.

#### *2.2.1. A typical day on Mars*

The day-to-day operations at the MDRS can be described as follows. The crew wakes up at 7:30. Then directly follows a twenty minute morning sport session, before having breakfast, which is typically the right moment for daily medical examinations.

Extra-vehicular activities (EVAs) always take place during the morning. The crew members that remain inside MDRS stay in permanent contact with the EVA party, while performing the daily chores. Scientific activities then take place every day from 1:30pm to 6pm. The crew members usually work on separate places, depending on their research field: the crew botanist stays in the green hab, biologists and chemists in the science dome, the astronomer in the observatory, engineers work in the RAM, *etc*.

All research activities are interrupted around 6pm, in order to prepare for the daily communication window (capcom), from 7 to 8pm. Finally, the crew members, one by one, use the RAMS system to report past operations and schedule future ones. Therefore, each crew member uses the *Romie* system *–and answer the questions and exercises defined in the scope of our study–* once a day, for approximately 30 to 60 minutes. The remaining of the evening constitutes a privileged, necessary moment for socialising.

**Fig. 2**. Panoramic pictures of some of the Mars Desert Research Station (MDRS) elements, from inside. From top to bottom: upper deck, lower deck, EVA preparation room, science dome, green hab.

#### *2.2.2. Time-eaters at MDRS*

Previous studies [17, 18, 19, 20, 21, 22, 23] have shown that there are many 'time-eaters' in a day at the MDRS during a simulated Mars stay mission. Typically, regarding the average unproductive time of the 76th rotation in 2009, a hypothetical average crew member would sleep an average of 8.5 hours, eat breakfast during an average of 44 minutes, lunch for 48 minutes and dinner for 57 minutes, spend an average of 3 hours doing chores and 1.5 hour doing maintenance, and spend an average of 1.5 hour on evening common activities, which sums up to 17 hours, leaving only approximately 7 hours for scientific work. Several recommendations were made to improve the design in order to optimize the traffic and to decrease the time spent unproductively from a scientific point of view. Yet, the "time-eaters" cannot be completely avoided. The system proposed in the current study comes in addition to these recommendations, as we investigate an AI based decision system to optimise productivity while leveraging unpredictable time deviations.

### 2.3. Research projects

Each analogue astronaut has her/his own research objectives for the mission. In fact, each astronaut prepared one different research project to be carried out at MDRS. There are thus eight research projects, from eight different fields such as biology, botanic, engineering, astronomy or medicine:

**Soil dielectric 3D map:** Using a ground penetrating radar, installed on a vehicle, the dielectric properties of the

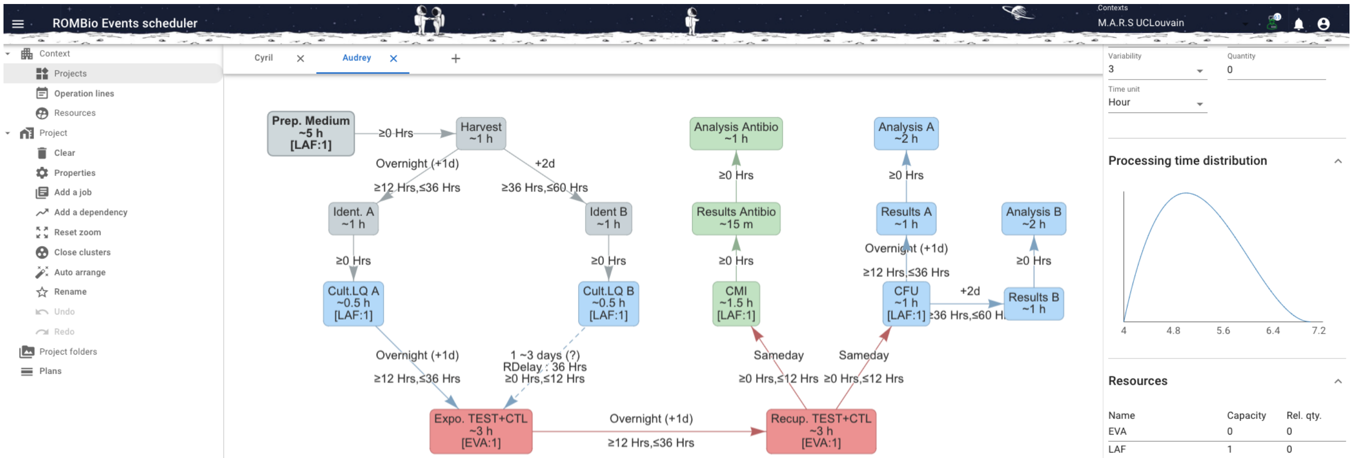


**Fig. 3**. Audrey's research project *"Survival of human flora bacteria"*. Amongst the activity properties in the right panel, we note that the selected activity *Prep. Medium* requires the *LAF* (laminar air flow) resource. The temporal constraint between *Cult.LQ B* and *Expo.TEST+CTL* involves a stochastic delay, between 1 and 3 sols (Martian days).

soil surrounding the station will be measured and projected on a 3D map, constructed by photogrammetry using a drone. Such a map could be exploited to optimize future irrigation systems. This project is lead by *Cyril Wain* (crew commander). **3D printing:** This experiment exploits 3D printing scaffolds in bio-ink to seed stem cells, and performs mechanical stress-strain tests on the resulting micro-architecture. This project is lead by *Ignacio Sanchez Casla* (crew astronomer). **Sleeping hypnosis:** This project tests an hypnosis technique, used in medicine before falling asleep, to help the astronauts having better, deeper sleeps. This project is lead by *Julien Meert* (crew engineer). **ExFix:** Accidents and injuries on Mars are dangerous. This project studies a low-cost external bones fixator, which remains accessible, fast and easily achievable by any astronaut without surgical training. This project is lead by *Julie Manon* (health & safety officer). **Metabolic changes:** The lower gravity of Mars, its environment and the nutrition changes will have a big impact on future crews' metabolisms. Here, a protocol is developed for the monitoring of essential parameters of the health and metabolism of the crew members. This project is lead by *Jean Jacobs* (executive officer). **Insects in the astronauts' diet:** Insects constitute a potential alternative food solution for astronaut crews. The viability and yield rate of three insect species (orthoptera, beetle and lepidopteran) are experimented under Martian conditions. This project is lead by *Sirga Drouet* (crew journalist). **Survival of human flora bacteria:** The survival of some human flora bacteria and the efficacy of several antibiotics under Martian environmental conditions is experimentally studied. This project is lead by *Audrey Comein* (crew scientist). **The effect of biofertilizers in a Martian soil substrate:** This experiment analyzes how a closed environment like the MDRS station and with a Martian regolith, the caloric intake of astronauts can be filled thanks to biofertilizers in small quantities. This project is lead by *Cheyenne Chamart* (greenhab officer).

### *2.3.1. Modelling and Scheduling on the RAMS system*

Figure 3 shows the modelling of one of the research projects, as encoded in *Romie*. From an operational point of view, this model has interesting properties. It involves a resource shared with other scientists: the laminar air flow (LAF). Since there is only one LAF in the station's science dome, this prevents other activities (belonging to other projects), also requiring the LAF, to be carried out at the same time. Another point of interest is the temporal constraint between *Cult.LQ B* and *Expo.TEST+CTL*, which involves a stochastic delay. In fact, the delay that must be waited between those two activities (1, 2 or 3 days) depends on the time needed by the bacteria to grow, and it is totally unpredictable by nature. Finally, there are temporal constraints, stating that some activity should not start sooner and/or later than a defined amount of time after some other activity.

The temporal constraints present in this model may potentially lead to a project failure, due to the underlying temporal uncertainty. Yet, another kind of complexity lies in models that involve the participation of several crew members, in addition to shared equipment. Figure 4 shows an example of an optimized *provisional* schedule, as obtained using *Romie*'s optimization engine, for all eight research project during the entire mission. In this schedule, the activities involved in Julie's research project *ExFix* are highlighted. We directly see that many of these require time within the schedule of the other crew members.

The RAMS system here not only allows to check the deterministic KPIs, but also some probabilistic ones. From a *deterministic* point of view, when all the durations are assumed

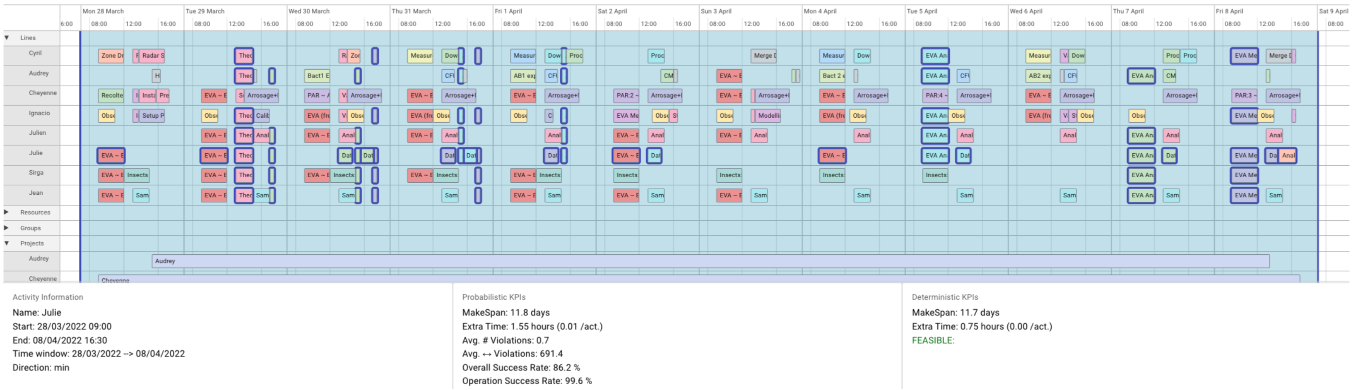


**Fig. 4**. A provisional schedule.

to require their nominal operational time, this planning is *feasible*. For example, in the project modelled in Figure 3, provided that the delay between *Cult.LQ B* and *Expo.TEST+CTL* will reveal to be exactly two sols. However, when taking *uncertainty* into account, then the mission probability of success is of 86.2%. Here the system only takes temporal uncertainty into account, not the fact that the activities themselves could be failed, requiring a rescheduling. Rescheduling operations, as well as adding new operations on the fly, will be part of the astronauts' daily manipulation on the system.

## 3. RESULTS

The astronauts were asked to evaluate their experience of the system and the quality of the computed decisions, before and at different stages during the mission. The a priori stage, before the mission, is called sol zero ($S_0$). A sol is a day on Mars. The subsequent stages are $S_4$, $S_8$ and $S_{12}$, for sols four (early mission), eight and twelve (end of the mission).

Participants were instructed to complete an UEQ+ questionnaire (User Experience Questionnaire) [24], a modular extension of the UEQ evaluation method in which we selected 12 scales to focus on evaluating the user experience of participants interacting with the system: ATTRACTIVENESS, EFFICIENCY, PERSPICUITY, DEPENDABILITY, STIMULATION, NOVELTY, TRUST, ADAPTABILITY, USEFULNESS, VISUAL AESTHETICS, INTUITIVE USE, and TRUSTWORTHINESS OF CONTENT. Each scale is in turn decomposed into four subscales or items to be evaluated (*e.g.* attractiveness is decomposed into four subscales: annoying vs. enjoyable, bad vs. good, unpleasant vs. pleasant, and unfriendly vs. friendly), each subscale being a differential scale with 7 points between items of each pair (annoying o o o o o o o enjoyable). We measure each item employing a 7-point Likert-type scale with response categories "Strongly disagree" (=1) to "Strongly agree" (=7).

UEQ+ was selected as an evaluation method because it is a modular and modern interface evaluation method where scales can be decided based on the interface to evaluate and covers the user experience (UX), not just usability, as assessed by questionnaires such as IBM PSSUQ [25]. UEQ+ is also easy to administer to participants and remains valid even with a limited number of participants. Participant answers are interpreted with the UEQ data analysis tool. According to [24], "it is extremely unlikely to observe values above +2 or below -2,..., the standard interpretation of the scale means is that values between -0.8 and 0.8 represent a neutral evaluation of the corresponding scale, values superior to 0.8 represent a positive evaluation, and values inferior to -0.08 represent a negative evaluation".

### 3.1. Before the mission: $S_0$

We collected demographic information from the participants at the beginning of the first session, at sol zero ($S_0$), that is *before the beginning of the mission*. Two scales are negatively assessed in the neutral zone. First, PERSPICUITY ($M$=−0.22, $SD$=1.36) expresses that the participants did not quickly familiarize themselves with the system, which could be explained that this was their first discovery of the system. Second, VISUAL AESTHETICS ($M$=−0.34, $SD$=1.63) was also negatively assessed. INTUITIVE USE ($M$=0.31, $SD$=1.45) is the only scale assessed positively in the neutral zone. However, three scales were borderline, that is, ATTRACTIVENESS ($M$=0.86, $SD$=0.88), ADAPTABILITY ($M$=0.88, $SD$=1.19), and TRUST ($M$=0.94, $SD$=1.27), thus reflecting that participants were still not convinced that the system fulfilled their expectations with respect to these three important aspects. Fortunately, six scales of 12 are positively assessed, even above the threshold, suggesting that the participants felt these aspects already well fulfilled at first glance: USEFULNESS ($M$=2.00, $SD$=1.15), STIMULATION ($M$=1.91, $SD$=1.07), TRUSTWORTHINESS ($M$=1.81, $SD$=0.98), NOVELTY ($M$=1.78, $SD$=0.96), DEPENDABILITY ($M$=1.38, $SD$=0.96), and EFFICIENCY ($M$=1.16, $SD$=1.09).

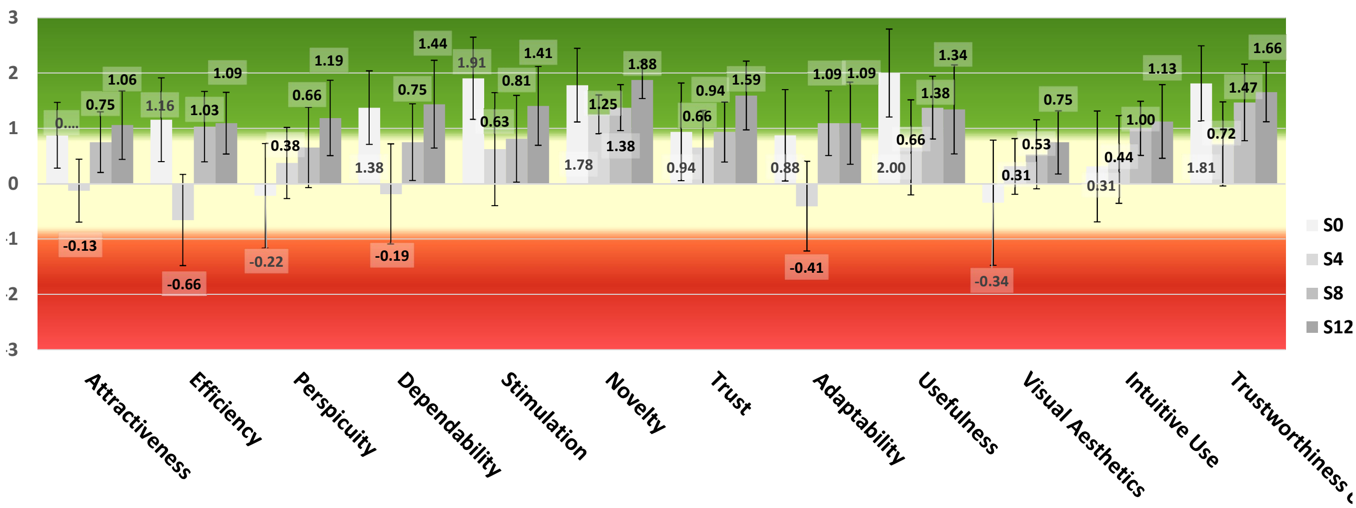


**Fig. 5**. UEQ+ Scale means for all sessions $S_0, S_4, S_8, S_{12}$. Error bars show a confidence interval of 95%.

### 3.2. During the mission: $S_4$, $S_8$, $S_{12}$

At an early stage of the mission, at the end of the fourth day, $\boldsymbol{S_4}$ (sol four), only Novelty ($M$=1.25, $SD$=0.50) exceeds the 0.8 threshold with small dispersion, thus meaning that participants all recognise that the software was novel, original for them, partly because they were never confronted to any similar software. Seven of 12 scales are positively assessed in the neutral interval, representing a slight improvement with respect to $S_0$: Trustworthiness of content ($M$=0.72, $SD$=1.10) Trust ($M$=0.66, $SD$=0.96), Usefulness ($M$=0.66, $SD$=1.24), Stimulation ($M$=0.63, $SD$=1.47), Intuitive Use ($M$=0.44, $SD$=1.14), Perspicuity ($M$=0.38, $SD$=0.93)), and Visual Aesthetics ($M$=0.31, $SD$=0.73). The three most positive scales refer to the utility character of the application, which is considered as the most important part first. Four scales are negatively assessed in the neutral interval, pointing to areas of improvement: Efficiency ($M$=−0.66, $SD$=1.19), Adaptability ($M$=−0.41, $SD$=1.17), Dependability ($M$=−0.19, $SD$=1.31), and Attractiveness ($M$=−0.13, $SD$=0.82).

For the first time, at $\boldsymbol{S_8}$ (sol eight) all scales are positively assessed with only four belonging to the neutral zone: Attractiveness ($M$=0.75, $SD$=0.79), Dependability ($M$=0.75, $SD$=1.00), Perspicuity ($M$=0.66, $SD$=1.05), and Visual aesthetics ($M$=0.53, $SD$=0.90). The remaining eight scales are located above the threshold as follows in decreasing order of their mean: Trustworthiness of content ($M$=1.47, $SD$=1.00), Usefulness ($M$=1.38, $SD$=0.82), Adaptability ($M$=1.09, $SD$=0.84), Novelty ($M$=1.38, $SD$=0.60), Efficiency ($M$=1.03, $SD$=0.92), Intuitive use ($M$=1.00, $SD$=0.71), Trust ($M$=0.94, $SD$=0.79), and Stimulation ($M$=0.81, $SD$=1.13). Trustworthiness of content received the highest result ($M$=1.75), suggesting that participants progressively acquire more trust in manipulating the data. The functions attached to these data are well perceived based on Usefulness.

Finally, session $\boldsymbol{S_{12}}$ (sol twelve) obtained all scale means above the threshold, indicating a very positive appreciation of the software, except for Visual Aesthetics ($M$=0.75, $SD$=0.83). Surprisingly, Novelty obtained the highest scale mean ($M$=1.88) with the smallest deviation ($SD$=0.48) of all scales, suggesting that participants concur to estimate that the software stays original, even after several usages. Trustworthiness of content remains the second highest scale means ($M$=1.66, $SD$=0.77) as it was the case before. Attractiveness ($M$=1.06, $SD$=0.90) suffered from the lowest mean, suggesting that this factor does not deteriorate much the overall software quality. Just before this factor, Adaptability ($M$=1.09, $SD$=1.07) and Efficiency ($M$=1.09, $SD$=0.80) share the second lowest scale mean.

### 3.3. Overall analysis

Most scales obtained a high mean for the first session $S_0$, which dramatically decreases for the first session $S_4$ carried out in real conditions, revealing a different appreciation of the software between the ideal conditions *in vitro* and the real conditions *in vivo*. Fortunately, these means positively evolve until reaching positive values above the threshold during the last session, representing a very nice evolution over time. These results suggest that participants, although they were probably influenced by the difficult conditions of the $S_4$ session, progressively improved their assessment, probably being less influenced by these conditions and more accustomed to deal with them. The results obtained for the last sessions $S_{12}$ therefore represent an overall stable assessment of the software after several continuous usages.

We observe that some scales largely improved since the beginning: PERSPICUITY received the best mean gain from one session to the last ($\Delta$=643%), followed by VISUAL AESTHETICS ($\Delta$=318%) and INTUITIVE USE ($\Delta$=260%, suggesting that the experience gained during the sessions positively impacted these scale means.

Four scales decreased between the first and the last session: USEFULNESS is reduced by $\Delta = -33\%$, followed by STIMULATION by $\Delta = -26\%$, TRUSTWORTHINESS OF CONTENT by $\Delta = -9\%$, and EFFICIENCY by $\Delta = -5\%$, suggesting that participants expressed their needs at a higher level of expectation during the first session than during the last one. This does not depreciate the overall quality of the software, but indicates that the experience accumulated by participants let them to adjust their assessment more precisely since all scale means in $S_{12}$ were highly positive. The progress acquired during successive sessions is therefore a determining factor for the adjustment of the scales in order to converge towards an equilibrium value representing a really stable value after a continuous interaction.

## 4. CONCLUSIONS AND FUTURE WORK

We study the capability of a crew of analogue astronauts, composed of novice planners, to manage the operational schedule of their mission in an autonomous setup, by using a computer-aided decision system. Techniques from human computer interaction (HCI) were exploited to measure and analyse how well the participants succeed at doing so: the astronauts were asked to evaluate their experience of the system and the quality of the computed decisions using an UEQ+ questionnaire.

The results gathered before, and at different stages of the mission, show that the proposed decision system appears as being an adequate approach, from a functional point of view (usefulness), whereas it is perceived as difficult to use by the participants, especially during the first days of the mission.

Empirical evidence has shown that even provided a strong provisional schedule, rethinking and reshaping all the a priori decisions related to the research projects, to be carried out during the mission, is unavoidable. As activities take place, the scientific objectives and constraints must be adapted according to unpredictable events. EVAs must be cancelled due to bad weather conditions. The entire project must be adapted to fit the limited duration of the mission. Due to the inherent complexity of the underlying combinatorial problem, modifying the schedule by hand is not an option. To that extent, the tested decision system includes an artificial intelligence, which proved its usefulness by computing optimised solutions to the scheduling problem, for the astronauts, based on a graphical description of their objectives and constraints. The main limitation of the approach lies in the learning time required by the participants to master the system. Future missions will need a more adequate preparation.

Future work will consider a more in-depth analysis of the UEQ+ results at two levels: (1) *intra-session*: Section 3.1 reported the evaluation before the mission in ideal, certainly unconstrained, experimental conditions while Section3.2 reported the evaluation of the the three sessions carried out during the mission. Beyond the 0.8 threshold, some scales are subject to a benchmarking which allows a more refined appreciation of their results. The benchmarking could be expanded to all scales and compared depending on the experimental conditions for a context-aware adaptation [26]. Furthermore, the importance levels rated by participants should be subject to a further analysis to map scales and the importance perceived by participants. A consistency reliability analysis should also be conducted to reveal how participants were reliable and consistent among them. (2) *inter-session*: an inferential statistical analysis should be performed to investigate whether some scales are assessed with a significant difference across sessions (e.g., between the first and the last real sessions) and/or are perceived significantly different from each other within a session. This analysis should be mitigated by importance rates.

## 5. REFERENCES


[1] Michael Saint-Guillain, Jonas Gibaszek, Tiago Stegun Vaquero, and Steve Chien, "Romie: A domain-independent tool for computer-aided robust operations management," *Engineering Applications of Artificial Intelligence*, vol. Volume 111, May 2022, 104801, 2022.

[2] Jessica J Marquez, Matthew J Miller, Tamar Cohen, Ivonne Deliz, David S Lees, Jimin Zheng, Yeon J Lee, Bob Kanefsky, Johannes Norheim, Matthew Deans, et al., "Future needs for science-driven geospatial and temporal extravehicular activity planning and execution," *Astrobiology*, vol. 19, no. 3, pp. 440–461, 2019.

[3] Ernest Bell and David Coan, "A review of the approach to iss increment crew eva training," in *AIAA SPACE 2007 Conference & Exposition*, 2012, p. 6236.

[4] Matthew J Miller, Kerry M McGuire, and Karen M Feigh, "Information flow model of human extravehicular activity operations," in *2015 IEEE Aerospace Conference*. IEEE, 2015, pp. 1–15.

[5] Andrew Mishkin, Young Lee, David Korth, and Troy LeBlanc, "Human-robotic missions to the moon and mars: operations design implications," in *2007 IEEE Aerospace Conference*. IEEE, 2007, pp. 1–10.

[6] Dean Eppler, Byron Adams, Doug Archer, Greg Baiden, Adrian Brown, William Carey, Barbara Cohen, Chris Condit, Cindy Evans, Corey Fortezzo, et al., "Desert research and technology studies (drats) 2010 science operations: operational approaches and lessons learned for

managing science during human planetary surface missions," *Acta Astronautica*, vol. 90, no. 2, pp. 224–241, 2013.

[7] Bret G Drake and D Watts Kevin, "Human exploration of mars design reference architecture 5.0, addendum# 2," 2014.

[8] David Applegate and William Cook, "A computational study of the job-shop scheduling problem," *ORSA Journal on computing*, vol. 3, no. 2, pp. 149–156, 1991.

[9] N G Hall and M J Magazine, "Maximizing the value of a space mission," *European Journal of Operational Research*, vol. 78, no. 2, pp. 224–241, 1994.

[10] Jessica J Marquez, Tamsyn Edwards, John A Karasinski, Candice N Lee, Megan C Shyr, Casey L Miller, and Summer L Brandt, "Human performance of novice schedulers for complex spaceflight operations timelines," *Human Factors*, p. 00187208211058913, 2021.

[11] S. A. Chien, G. Rabideau, D. Q. Tran, M. Troesch, F. Nespoli, M. Perez-Ayucar, M. Costa-Sitja, C. Vallat, B. Geiger, F. Vallejo, R. Andres, N. Altobelli, and M. Kueppers, "Activity-based scheduling of science campaigns for the rosetta orbiter," *Journal of Aerospace Information Systems*, vol. 18:10, pp. 711–727, 2021.

[12] Michael Saint-Guillain, Tiago Vaquero, Steve Chien, Jagriti Agrawal, and Jordan Abrahams, "Probabilistic temporal networks with ordinary distributions: Theory, robustness and expected utility," *Journal of Artificial Intelligence Research*, vol. 71, pp. 1091–1136, 2021.

[13] Michael Saint-Guillain, "Robust operations management on mars," in *Proceedings of the International Conference on Automated Planning and Scheduling*, 2019, vol. 29, pp. 368–376.

[14] Michael Saint-Guillain, Michael Brion, Eric Pauly, and Isabelle Sansdrap, "Robust advanced modelling and scheduling system (rams): From space exploration to real-world biomanufacturing," *Journal of Biotechnology (under publication)*, 2022.

[15] International Standard Organization, "ISO/IEC 9421-210:2019, ergonomics of human-system interaction — part 210: Human-centred design for interactive systems," 2019.

[16] Arnold P. O. S. Vermeeren, Effie Lai-Chong Law, Virpi Roto, Marianna Obrist, Jettie Hoonhout, and Kaisa Väänänen-Vainio-Mattila, "User experience evaluation methods: Current state and development needs," in *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, New York, NY, USA, 2010, NordiCHI '10, p. 521–530, Association for Computing Machinery.

[17] V Pletser, L Boche-Sauvan, BH Foing, and Eurogeomars Team, "European contribution to human aspect field investigation for future planetary habitat definition studies: field tests at mdrs on crew time utilization and habitat interfaces," in *ELGRA Biannual Symposium 'In the Footsteps of Columbus', Bonn, Germany*, 2009, vol. 26, p. 64.

[18] L. Boche-Sauvan, Vladimir Pletser, B Foing, and Eurogeomars Crew, "Human aspects study through industrial methods during an mdrs mission," in *NASA Lunar Science Forum, NASA Ames Conference Centre*, 2009.

[19] L Boche-Sauvan, V Pletser, BH Foing, and Eurogeomars Team, "Human aspects and habitat studies from eurogeomars campaign," in *EGU General Assembly Conference Abstracts*, 2009, p. 13323.

[20] Vladimir Pletser, "Crew time utilisation and habitat interface investigations for future planetary habitat definition studies: field tests at mdrs," *38th COSPAR Scientific Assembly*, vol. 38, pp. 2, 2010.

[21] Vladimir Pletser, "A mars human habitat: Recommendations on crew time utilisation and habitat interfaces," *Journal of Cosmology, Special Issue on 'The Human Mission to Mars: Colonizing the Red Planet'*, vol. 12, pp. 3928–3945, 2010.

[22] Cora S Thiel, Vladimir Pletser, and Bernard Foing, "Human crew-related aspects for astrobiology research," *International Journal of Astrobiology*, vol. 10, no. 3, pp. 255–267, 2011.

[23] Vladimir Pletser and Bernard Foing, "European contribution to human aspect investigations for future planetary habitat definition studies: Field tests at mdrs on crew time utilisation and habitat interfaces," *Microgravity Science and Technology*, vol. 23, no. 2, pp. 199–214, 2011.

[24] Martin Schrepp and Jörg Thomaschewski, "Design and validation of a framework for the creation of user experience questionnaires," *Int. J. Interact. Multim. Artif. Intell.*, vol. 5, no. 7, pp. 88–95, 2019.

[25] James R. Lewis, "Sample sizes for usability tests: Mostly math, not magic," *Interactions*, vol. 13, no. 6, pp. 29–33, Nov. 2006.

[26] Vivian G. Motti and Jean Vanderdonckt, "A computational framework for context-aware adaptation of user interfaces," in *Proc. of the 7th IEEE Int. Conf. on Research Challenges in Information Science*, May 2013, RCIS '13, pp. 1–12.