The Role of Trust and Usability in Enabling Spaceflight Crew Autonomy

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

Dakota Sullivan[2]
Universities Space Research Association, Columbia, MD, 21046, USA

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

Future long duration exploration missions will require an increased use of onboard automated systems as spaceflight crews venture further than before and have longer communications delays with ground support. The design of these systems must support appropriate crew trust and have sufficient usability to enable spaceflight crew autonomy or risk being misused while crews wait to communicate with ground support. We evaluated trust & usability in our self-scheduling tool, Playbook, for crew mission timelines. Data was collected in a controlled lab experiment with 31 participants. Participants in the study conducted two tasks: scheduling, where participants were responsible for scheduling a majority of a day's operational tasks, and rescheduling, where participants were provided a schedule and asked to reschedule higher priority activities. We found a significant correlation between system trust and usability, irrespective of self-scheduling tasks. We conclude that system usability may play a bigger role in how trust is learned while conducting novel crew autonomy tasks such as self-scheduling. Future research should investigate the role of usability to encourage appropriate trust in onboard automated crew systems and enable crew autonomy.

Nomenclature

HCAAM = Human Capabilities Assessments for Autonomous Missions
LDEM = Long Duration Exploration Mission
MCC = Mission Control Center
TLX = Task Load Index
TAS = Trust in Automated Systems
UEQ = User Experience Questionnaire
VNSCOR = Virtual NASA Specialized Center of Research

I. Introduction

As NASA considers future human long duration exploration missions (LDEMs), there is an increased emphasis on developing systems that enable crew autonomy. Astronauts will have to operate more autonomously from Mission Control Center (MCC) due to longer communication delays and limited bandwidth to ground personnel.

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

Future crew systems will integrate new technologies and capabilities previously delegated to MCC to enable astronauts to perform operational tasks more autonomously. Design of these systems should consider various aspects of human spaceflight operations: they must support safety-critical operations by allowing astronauts to work efficiently and effectively; they must help crew adequately develop and maintain situation awareness without excessively adding to workload; and they must permit suitable trust calibration without causing frustration. NASA is studying these types of systems through the Virtual NASA Specialized Center of Research (VNSCOR) for Human Capabilities Assessments for Autonomous Missions (HCAAM). As part of the VNSCOR, our team investigates one such technology, a software tool that supports astronaut self-scheduling. This paper specifically considers the impact of trust on self-scheduling as it relates to various factors such as workload, usability, and participants' expertise.

II. Background

Trust is the concept describing that "an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability" [1]. Recently, there has been much research focused on understanding and measuring the trust that people have in automated or software systems. The concept of trust is an important design consideration because it influences the overall effectiveness of collaboration between automated systems and their users [2]. For instance, a system that is not considered trustworthy is not likely to be used, while excessive trust may instead lead to overreliance. Within the domain of human spaceflight, appropriately calibrating astronauts' trust in a system is imperative for safe missions. Astronauts rely on various automated systems that are essential to mission operations. Future exploration missions will see an increase in use of automated robotics and software systems to accomplish a diverse, complex set of tasks. System trust is a continued area of interest for NASA as more focus is placed on deep space missions [3].

Our research regarding trust is focused on understanding the impact it has on self-scheduling. Self-scheduling is part of a new concept of operations that will enable crew autonomy in future exploration spaceflight missions. Currently, astronauts are not allowed to manage their own schedule in space; instead, they follow a predetermined schedule created by a large team of flight controllers. If an astronaut falls behind or gets ahead while following the schedule then they coordinate with flight controllers to see how to rearrange their schedule to accommodate for the change. Future missions will have to tackle long, intermittent communication transmission latencies between Earth and space, resulting in idle time as astronauts wait for feedback from flight controllers. Allowing astronauts to conduct self-scheduling (i.e., schedule or reschedule assigned activities on their own) is expected to increase crew autonomy and permits astronauts to work more efficiently, maximizing their limited time.

This new concept of operations will require astronauts to manipulate their own schedule using new software tools. Astronauts will also have to make sure that their new schedules abide by the multiple constraints and limited resources available in spaceflight. If NASA wants astronauts to conduct self-scheduling, then the crew will need to do so safely, making sure that the new schedules are feasible to execute. Self-scheduling software tools need to not only support the task but also foment trust among novice planners. Thus, we set out to explore for the first time if trust is elicited by our scheduling software tool, Playbook. Our approach centered around the three-layered framework for conceptualizing trust variability proposed by Hoff & Bashir [4]: dispositional trust, situational trust, and learned trust. As such, we specifically explored each type of trust, correlating overall trust measures with other covariates.

III. Methodology

We conducted a human-in-the-loop experiment where participants remotely completed one of two tasks: scheduling or rescheduling a spaceflight-like crew timeline. Thirty-one participants were recruited and trained to use the self-scheduling software, Playbook [5, 6]. Playbook permits users to select activities from a list, from which then they can drag and drop them into a timeline. Once scheduled, Playbook checks if an activity's constraints are satisfied, and if not, Playbook provides visual indicators that the activity is in violation (see Fig. 1).

Participants were assigned to groups to complete one of the two self-scheduling tasks. Participants in the scheduling group were asked to schedule 24 activities into a relatively empty schedule. Activities were grouped into high, medium, and low priority. The rescheduling group required participants to manipulate those same 24 activities, most of which were initially scheduled on the timeline. Participants were instructed to ensure that all the high priority activities were scheduled, then schedule the remaining activities in order of priority. Participants in both groups were instructed that none of the scheduled activities could be left in violation, and both groups had to complete nine trials. See Marquez et al. [7] for a complete description of the experiment and performance results.

Fig. 1 Playbook's timeline view showing two activities, highlighted in red, which are currently in violation.

Before the experiment, participants were asked to fill out a demographics survey, including information about their planning and scheduling experience, confidence with technology (similar to Correia et al. [8]), and dispositional trust [9]. Scheduling experience was rated on an 8-point scale where 0 indicates none and 7 indicates greater than ten years. The confidence with technology survey asked participants to rate their degree of confidence from 0 ("cannot do it at all") to 100 ("highly certain they can do it") on fifteen statements concerning touchscreen technologies. This survey was deployed as, in a previous study, a participant who had no experience with touchscreen technologies did not want to use the self-scheduling tool. The dispositional trust survey includes twenty statements where participants determine if they agree or disagree on a five-point scale. After each self-scheduling trial, participants completed the NASA-Task Load Index (TLX) [10], a questionnaire used to assess subjective workload. The NASA-TLX accounts for six metrics including mental, physical, and temporal demand as well as performance, effort, and frustration. These metrics were aggregated to calculate a single workload score for each trial.

Once the self-scheduling trials were completed, participants completed a trust survey using the Trust in Automated Systems (TAS) Scale [11] and the User Experience Questionnaire (UEQ) [12]. The TAS is a 12-item scale incorporating positive and negative dimensions of trust as well as a user's familiarity with a given automated system; these were aggregated to calculate an individual trust score. The UEQ is a 26-point scale that measures six metrics such as attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Five of these metrics can be aggregated to create a Pragmatic Score (perspicuity, efficiency, and dependability) and a Hedonic Score (stimulation and novelty).

IV. Results

Overall, participant perception of trust in self-scheduling software was rated highly. The average trust score across all participants was rated 5.94 (SD = 0.64) on a 7-point scale (see Fig. 2 for distribution). There was minimal difference in trust between the type of task according to Welch's t-test, $t(27.58) = 0.08$, $p = 0.93$. Scheduling participants rated trust at 5.95 on average (SD = 0.65) and rescheduling participants rated trust at 5.93 (SD = 0.66). Irrespective of task type, across negative trust components (i.e., deceptiveness, underhanded behavior, harmful outcomes, etc.) participants rated self-scheduling software at 1.27 on average (SD = 0.69) while rating positive components (i.e., secure, dependable, reliable, etc.) at 5.40 (SD = 1.22). These findings show that participants generally found the self-scheduling software to be trustworthy, with very minimal negative perception and high positive perception.

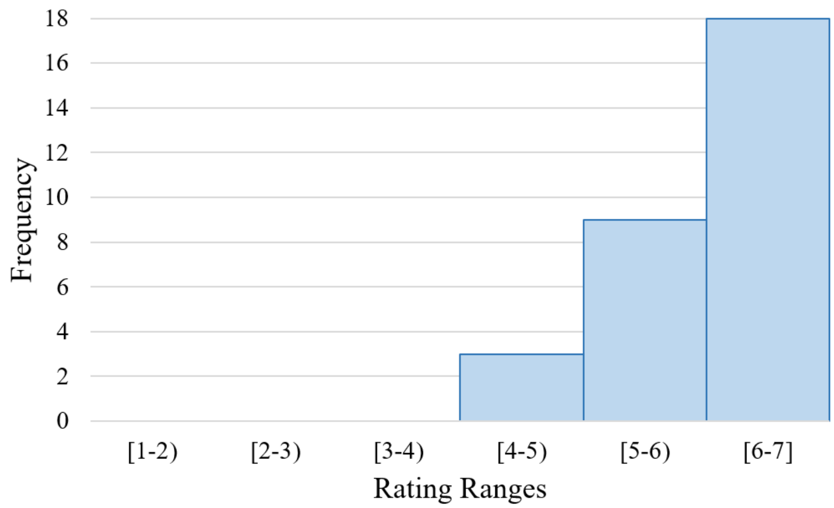

Fig. 2 Distribution of trust scores (TAS) in Playbook for all participants.

Dispositional Trust: We next examined whether trust of the self-scheduling software was based on the participants' predisposition to trust. Schaefer et al. [2] proposed a model of the various elements that impact user trust, which includes human-related elements (i.e., user traits, states, cognitive factors, and emotive factors), partner-related elements (i.e., features and capabilities), and environment-related elements (i.e., team collaboration and task context). We specifically examined if trust was affected by the participants' dispositional trust and participant's confidence with technology. After coding and inverting negative responses, the average dispositional trust score across all participants was 3.84 (SD = 0.58) on a 5-point scale. Confidence in technology was rated highly by participants, with an average score of 86.79 (SD = 8.60) on a 100-point scale. We conducted correlation tests (Pearson or Kendall rank, depending on whether the data was normally distributed) and there were no significant correlations between trust and the two different covariates. This suggests that dispositional trust was not a factor in determining overall trust for the controlled experiment.

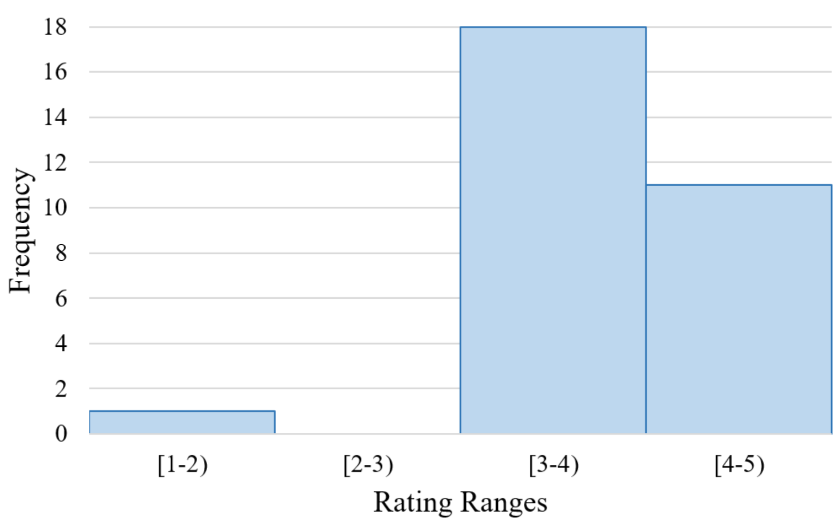

Fig. 3 Distribution of dispositional trust scores for all participants.

Situational Trust: Hoff & Bashir [4] postulate two sources of variability in situational trust: external environment factors (e.g., task difficulty, workload) and internal, context-dependent operator characteristics (e.g., mood, subject matter expertise). We examined task workload [7] and participants' reported scheduling expertise (see Fig. 4 for distribution) as potential proxies for external and internal factors influencing situational trust. The average planning and scheduling experience rating for participants was 1.60 (SD = 2.24), indicating they had limited experience with planning and scheduling. The average NASA-TLX workload score was 44.21 (SD = 19.53). There were no significant correlations between trust and workload nor trust and years of planning/scheduling experience, suggesting that situational trust was also not a factor in determining overall trust for the controlled experiment. There is likely insufficient variability in task workload and planning/scheduling experience compared to trust scores to make strong claims about this relationship.

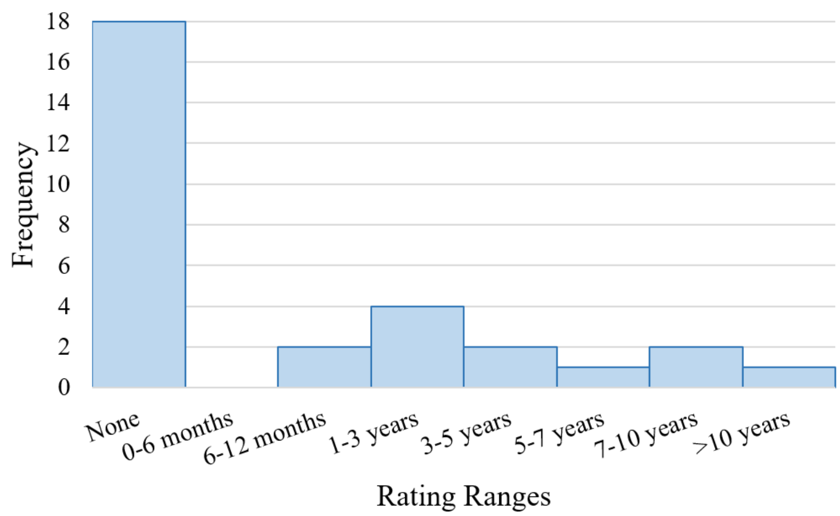

Fig. 4 Distribution of participants' reported scheduling experience.

Learned Trust: Learned trust, as the name implies, develops from a user's past or current interactions with a system [14]. As our participants had no previous experience with the self-scheduling software system Playbook, learned trust is sourced from current interactions with the system. Throughout the trials, the system performance was constant (i.e., it was consistently reliable and predictable), and as such, the system's design features (e.g., appearance, ease-of-use) would play a role in learned trust. We identified a strong, significant correlation between trust and usability.

Usability was measured using the UEQ survey, and detailed UEQ results are summarized in Shelat et al. [13]. UEQ scores could range from -3 to +3, with values between -0.8 and 0.8 representing a neutral evaluation, values greater than 0.8 representing a positive evaluation, and values less than -0.8 representing a negative evaluation. Overall, the self-scheduling software was evaluated positively on most UEQ metrics. Welch's t-tests were used to compare UEQ ratings between scheduling and rescheduling groups, and there were no significant results between task types. This suggests that task type was not a factor in our experiment regarding perceptions of usability in self-scheduling software, which is consistent with our findings regarding overall trust. There were, however, significant correlations between trust and several of the UEQ metrics. There was a strong correlation between trust and attractiveness, $r(27) = 0.69$, $p < 0.001$; efficiency, $r(27) = 0.64$, $p < 0.001$; dependability, $r(27) = 0.67$, $p < 0.001$; stimulation, $r(27) = 0.59$, $p < 0.001$; and novelty, $r(27) = 0.51$, $p < 0.01$. There was a moderate correlation between trust and perspicuity, $r(27) = 0.41$, $p < 0.01$. These results are surprising as there is limited evidence to suggest that there is a strong relationship between system trust and its usability.

V. Discussion and Conclusions

In order to better understand trust in the context of self-scheduling, we explored correlations between trust and factors associated with dispositional, situational, and learned trust. Our results suggest that learned trust is a main factor that influences trust in the self-scheduling software tool Playbook. More specifically, we identified a strong relationship between usability and participants' trust.

Previous studies have analyzed the impact of usability on trust within contexts such as online games [14], wearable fitness devices [15], website loyalty [16], mobile commerce [17, 18] consumer products, and voting systems [19]. One study suggests that factors such as advertising or brand reputation may obscure the connection between usability and trust as the user does not depend exclusively on their experience with the system to develop trust [19]. However, in the absence of these extraneous factors, user experience and perception of system usability are critical. Within the context of spaceflight, systems and tools are specialized for given tasks and are without alternative options. Crew trust is therefore based only upon experience with the system rather than branding or choreographed representations of that system. Additional studies on trust and usability have been conducted within the context of spaceflight [20-24], however, we have not identified any study that directly explored the relationship between these factors. Increased usability may be a novel method of improving trust within this domain and warrants further exploration.

Our research suggests that system usability may play a bigger role in how trust is learned while conducting novel crew autonomy tasks such as self-scheduling. Playbook's development has centered around following human-computer interaction principles, emphasizing user-centered design. Many usability evaluations in analogous environments [6, 25, 26] have led to a software tool that is considered easy-to-use [13] with a low-barrier of entry for astronauts [27, 28]. An unexpected benefit for emphasizing usability in self-scheduling software has been a trustworthy tool for future astronauts.

Future software aids designed to improve self-scheduling performance should prioritize features that further improve usability as it promotes and maintains user trust. This is particularly relevant as our team considers mixed-initiative scheduling [29], which leverages more automated methods for planning and scheduling. Increasing the amount of automation in software tools could lead to automation bias, increase incidence of errors, and resultantly diminish user trust [30]. One potential way of calibrating user trust could be through the design of features that are highly user-friendly and have positive usability ratings. Furthermore, future systems that support crew autonomy should consider both trust and usability as enabling design characteristics.

Acknowledgements

This research was funded by the NASA Human Research Program's Human Factors and Behavior Performance Element (NASA Program Announcement number 80JSC017N0001-BPBA) Human Capabilities Assessment for Autonomous Missions (HCAAM) Virtual NASA Specialized Center of Research (VNSCOR) effort. The authors would like to thank Tamsyn Edwards, Megan Shyr, Candice Lee, and Casey Miller for their efforts in conducting this research experiment.

References

[1] Lee, J. D., and See, K. A. "Trust in Automation: Designing for Appropriate Reliance." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 46, No. 1, 2004, pp. 50–80.
doi: 10.1518/hfes.46.1.50.30392

[2] Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., and Hancock, P. A. "A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 58, No. 3, 2016, pp. 377–400.
doi: 10.1177/0018720816634228

[3] Karasinski, J., Holder, S., Robinson, S., and Marquez, J. "Deep Space Human-Systems Research Recommendations for Future Human-Automation/Robotic Integration." 2020. https://ntrs.nasa.gov/citations/20205004361

[4] Hoff, K. A., and Bashir, M. "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 57, No. 3, 2015, pp. 407–434.
doi: 10.1177/0018720814547570

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

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

[7] Marquez, J. J., Edwards, T., Karasinski, J. A., Lee, C. N., Shyr, M. C., Miller, C. L., and Brandt, S. L. "Human Performance of Novice Schedulers for Complex Spaceflight Operations Timelines." *Human Factors: The Journal of the Human Factors and*

Ergonomics Society, 2021, p. 001872082110589.
doi: 10.1177/00187208211058913
[8] Correia, J., Compeau, D., and Thatcher, J. “Implications of Technological Progress for the Measurement of Technology Acceptance Variables: The Case of Self-Efficacy.” *DIGIT 2016 Proceedings*, 2016. https://aisel.aisnet.org/digit2016/6
[9] Singh, I. L., Molloy, R., and Parasuraman, R. “Automation- Induced ‘Complacency’: Development of the Complacency-Potential Rating Scale.” *The International Journal of Aviation Psychology*, Vol. 3, No. 2, 1993, pp. 111–122.
doi: 10.1207/s15327108ijap0302_2
[10] Hart, S. G., and Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*, Elsevier, 1988, pp. 139–183.
doi: 10.1016/S0166-4115(08)62386-9
[11] Jian, J.-Y., Bisantz, A. M., Drury, C. G., and Llinas, J. *Foundations for an Empirically Determined Scale of Trust in Automated Systems:* Defense Technical Information Center, Fort Belvoir, VA, 1998.
doi: 10.21236/ADA388787
[12] Schrepp, M., Hinderks, A., and Thomaschewski, J. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience* (A. Marcus, ed.), Springer International Publishing, Cham, 2014, pp. 383–392.
doi: 10.1007/978-3-319-07668-3_37
[13] Shelat, S., Karasinski, J. A., Flynn-Evans, E. E., and Marquez, J. J., “Evaluation of User Experience of Self-scheduling Software for Astronauts: Defining a Satisfaction Baseline.” In *Engineering Psychology and Cognitive Ergonomics* (D. Harris and W.-C. Li, eds.), Springer International Publishing, Cham, 2022, pp. 433–445.
doi: 10.1007/978-3-031-06086-1_34
[14] Gao, Y. “Factors Influencing User Trust in Online Games.” *The Electronic Library*, Vol. 23, No. 5, 2005, pp. 533–538.
doi: 10.1108/02640470510631245
[15] Rupp, M. A., Michaelis, J. R., McConnell, D. S., and Smither, J. A. “The Role of Individual Differences on Perceptions of Wearable Fitness Device Trust, Usability, and Motivational Impact.” *Applied Ergonomics*, Vol. 70, 2018, pp. 77–87.
doi: 10.1016/j.apergo.2018.02.005
[16] Flavián, C., Guinalíu, M., and Gurrea, R. “The Role Played by Perceived Usability, Satisfaction and Consumer Trust on Website Loyalty.” *Information & Management*, Vol. 43, No. 1, 2006, pp. 1–14.
doi: 10.1016/j.im.2005.01.002
[17] Sarkar, S., Chauhan, S., and Khare, A. “A Meta-Analysis of Antecedents and Consequences of Trust in Mobile Commerce.” *International Journal of Information Management*, Vol. 50, 2020, pp. 286–301.
doi: 10.1016/j.ijinfomgt.2019.08.008
[18] Li, Y.-M., and Yeh, Y.-S. “Increasing Trust in Mobile Commerce through Design Aesthetics.” *Computers in Human Behavior*, Vol. 26, No. 4, 2010, pp. 673–684.
doi: 10.1016/j.chb.2010.01.004
[19] Acemyan, C. Z., and Kortum, P. “The Relationship Between Trust and Usability in Systems.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56, No. 1, 2012, pp. 1842–1846.
doi: 10.1177/1071181312561371
[20] Woods, A., Iwig, C., Dinh, J., and Salas, E. “Informing the Development of a Safety and Performance Metric Selection Toolkit: Subject Matter Experts Weigh In.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60, No. 1, 2016, pp. 1354–1358.
doi: 10.1177/1541931213601312
[21] Yashar, M., Marquez, J., Menon, J., and Torron, I. Preliminary Design of an ‘Autonomous Medical Response Agent’ Interface Prototype for Long-Duration Spaceflight. In *HCI International 2020 - Late Breaking Papers: User Experience Design and Case Studies* (C. Stephanidis, A. Marcus, E. Rosenzweig, P.-L. P. Rau, A. Moallem, and M. Rauterberg, eds.), Springer International Publishing, Cham, 2020, pp. 543–562.
doi: 10.1007/978-3-030-60114-0_37
[22] Widder, D. G., Dabbish, L., Herbsleb, J. D., Holloway, A., and Davidoff, S. “Trust in Collaborative Automation in High Stakes Software Engineering Work: A Case Study at NASA.” *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.
doi: 10.1145/3411764.3445650
[23] Rogers, H., Khasawneh, A., Bertrand, J., and Madathil, K. C. “An Investigation of the Effect of Latency on the Operator’s Trust and Performance for Manual Multi-Robot Teleoperated Tasks.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61, No. 1, 2017, pp. 390–394.
doi: 10.1177/1541931213601579
[24] Hollaway, D., Taylor, E., and Badger, J. “When the Eyes Don’t Have It: Autonomous Control of Deep Space Vehicles for Human Spaceflight.” In *ASCEND 2020*, American Institute of Aeronautics and Astronautics, 2020.
doi: 10.2514/6.2020-4163
[25] Hillenius, S. R., Marquez, J., Deliz, I., Kanefsky, B., Korth, D., Healy, M., Gibson, S., and Zheng, J. “Designing and Building a Crew-Centric Mobile Scheduling and Planning Tool for Exploring Crew Autonomy Concepts Onboard the International Space Station.” *International Space Station Research & Development Conference (ISSR&D 2016)*, 2016. https://ntrs.nasa.gov/citations/20190018055

[26] Marquez, J. J., Hillenius, S., Healy, M., and Silva-Martinez, J. “Lessons Learned from International Space Station Crew Autonomous Scheduling Test.” *International Workshop on Planning and Scheduling for Space (IWPSS 2019)*, 2019. No. ARC-E-DAA-TN70121
[27] Marquez, J. J., Hillenius, S., and Healy, M. “Increasing Human Spaceflight Capabilities: Demonstration of Crew Autonomy Through Self-Scheduling Onboard International Space Station.” *International Space Station Research & Development Conference (ISS R&D 2018)*, 2018. https://ntrs.nasa.gov/citations/20180005211
[28] Marquez, J. J., Hillenius, S., Zheng, J., Deliz, I., Kanefsky, B., and Gale, J. “Designing for Astronaut-Centric Planning and Scheduling Aids.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63, No. 1, 2019, pp. 468–469.
doi: 10.1177/1071181319631386
[29] Bresina, J. L., and Morris, P. H. “Mixed-Initiative Planning in Space Mission Operations.” *AI Magazine*, Vol. 28, No. 2, 2007, pp. 75–75.
doi: 10.1609/aimag.v28i2.2041
[30] Goddard, K., Roudsari, A., and Wyatt, J. C. “Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators.” *Journal of the American Medical Informatics Association*, Vol. 19, No. 1, 2012, pp. 121–127.
doi: 10.1136/amiajnl-2011-000089