Acta Astronautica 161 (2019) 108–114

Contents lists available at ScienceDirect

Acta Astronautica

journal homepage: www.elsevier.com/locate/actaastro

Team performance in space crews: Houston, we have a teamwork problem

Lindsay Larson[a,*], Harrison Wojcik[b], Ilya Gokhman[a], Leslie DeChurch[a], Suzanne Bell[b], Noshir Contractor[a]

[a] Department of Communication, Northwestern University, 2040 Campus Drive, Evanston, IL, 60201, USA
[b] Department of Psychology, DePaul University, 2219 N Kenmore Ave, Chicago, IL, 60614, USA

Keywords:
Team performance
Isolation
Confinement
Space simulation
Team process

A B S T R A C T

Space crews venturing beyond low Earth orbit will experience unprecedented levels of autonomy and unpredictable challenges. Mission success will require effective teamwork. How do teamwork capabilities change over time in isolation and confinement? To explore this question, 4, 4-member crews who participated in the 30-day campaign of the National Aeronautics and Space Agency (NASA)'s Human Exploration Research Analog (HERA) were observed. Crews endured isolation, confinement, and communication delays. Teamwork capabilities were observed along four critical dimensions: generate (creativity tasks), choose (intellective tasks), negotiate (cognitive conflict tasks), and execute (psychomotor tasks). A battery of team task was administered requiring the crew to generate, to choose, and to negotiate. Execute performance was assessed using NASA's multi-mission space exploration vehicle (MMSEV) task. Team task batteries were performed on Mission Days 11, 16, and 30. Execute performance was assessed on 18 of 30 days. Findings show behavioral team performance (cognitive conflict and psychomotor tasks) increases over time, whereas conceptual team performance (creativity and intellective tasks) declines. Implications of these results were considered for future research and the design of countermeasures that support crew functioning.

1. Introduction

Future space exploration missions will require extreme teamwork among crew members. As space crews venture farther from Earth, communication with mission control will become more difficult, with lag times of up to 20 min each way. This will require crews to work more autonomously than they have in prior missions. This degree of autonomy requires that all requisite expertise must be on board the spacecraft, meaning crews will be diverse in their disciplinary, technical, and cultural characteristics. These diverse crews will then need to rely on one another to perform a wide variety of complex tasks.

The larger the distance between the astronaut crew and mission control, the longer the communication delay between the two. For example, when the JPL [1] is communicating movements to an unmanned rover on the surface of Mars, there is about a 20-min delay between when JPL sends a command and when JPL receives confirmation from the rover that the command was received due to the distance the signal must travel. Astronaut crews will face similar delays. As future manned space exploration missions, like the mission to Mars, dare to go farther than ever before, crews will also likely encounter more unknowns. Although space agencies will work toward minimizing the risks associated with space flight, it will be impossible to plan for every contingency. Crew members will have to work together to address whatever unknowns arise during these unprecedented missions. Effective team performance will be essential to mission success.

Not only will space crews need to operate as autonomous, high performance teams for an extended period of time, but they will do so while facing environmental conditions that are distinct from many other kinds of teams: living and working together in an isolated and confined environment for an extended period of time. Understanding team performance under these conditions is a necessary first step in providing guidance for training and other interventions to sustain high-levels of crew performance over the entirety of a long-distance space exploration mission.

A team task battery was developed that assessed team performance on multiple dimensions important to team performance in general, and

Abbreviations: Alternative Uses Task, AUT; Extra-vehicular activity, EVA; Human Exploration Research Analog, HERA; International Space Station, ISS; Isolated and Confined Environment, ICE; Jet Propulsion Lab, JPL; Multi-mission space exploration vehicle, MMSEV; National Aeronautics and Space Agency, NASA

* Corresponding author.
E-mail addresses: lindsaylarson2020@u.northwestern.edu (L. Larson), wojci028@gmail.com (H. Wojcik), igokhman@u.northwestern.edu (I. Gokhman), dechurch@northwestern.edu (L. DeChurch), SBELL11@depaul.edu (S. Bell), nosh@northwestern.edu (N. Contractor).

https://doi.org/10.1016/j.actaastro.2019.04.052
Received 24 February 2019; Received in revised form 26 April 2019; Accepted 28 April 2019
Available online 03 May 2019

in long distance space exploration missions in particular. These include: creativity, problem solving, and ethical decision making. Data was also collected as part of the broader data collection campaign on psychomotor performance and analyzed in the analyses. Crew performance was observed on each type of task and dimension at three time points during each of four 30-day analog space missions.

1.1. Previous studies

In a research report published by researchers at NASA, team performance was identified as critical for future long-distance space exploration success [2]. Team performance is defined as the extent to which a team achieves its goals [3]. Space crews will work as a team to complete a variety of tasks. For example, a space crew must be able to creatively troubleshoot vehicle maintenance problems, complete onboard experimental studies, and make on-the-fly resource decisions, in addition to the crew's pre-scheduled day to day activities. These tasks will require crews to come together to think in creative and novel ways, effectively pool knowledge resources, and make morally challenging decisions to approach the unknowns that crews will inevitably encounter in the future of long-distance space exploration.

Yet the measurement of crew performance in space teams can be difficult. Performance in more regimented team environments, such as space or military teams, is often measured by whether or not an error has been made. However, when studying the performance of high achieving individuals, such as those selected to be astronauts, research has noted that errors are typically not critical until multiple errors have added up to a major failure [2,4]. Errors are also low occurrence. Therefore, NASA advocates for a focus on optimizing crew performance rather than minimizing errors [2]. This shift in focus better aligns with successfully completing the mission rather than just trying to avoid making errors at all costs with little to no consideration of achieving anything greater. Furthermore, the type of team performance dimensions most critical to long distance missions are more psycho-social in nature, involving creativity to solve unanticipated challenges, problem solving where members have different levels of expertise, and ethical decision-making emanating from the uncharted nature of a space mission.

One major challenge a long-distance space exploration crew will face is high levels of mission autonomy from mission control. In the past, space missions have relied extensively on direction from a centralized mission control with a plethora of experts to address any issue or question that might arise in mission. However, with the impending communication delay of longer distance space flight, crews will face a level of autonomy unprecedented in space travel, reminiscent of Christopher Columbus' voyage to the new world or Ernest Shackleton's journey to the South Pole. Once these journeys were planned and the crews set off, there was little support or intervention that could be provided.

A review of the team composition research in analog environments reveals that team composition will likely influence mission performance by affecting social integration and the processes and emergent states of the crew [5]. Past research [6] has demonstrated the differences and potential challenges of differing levels of autonomy in a simulated Mars mission. They manipulated crew autonomy levels (high vs. low) over the course of 105 days in a crew of 6 men in a Mars mission simulator in Moscow, Russia. For the first 10 weeks of the simulation, crew members underwent the low autonomy condition in which they were able to communicate with and receive daily instructions from mission control. For the final 5 weeks of the simulation, crew members underwent a high autonomy condition in which the crew had primary control over their schedule and a communication delay between the crew and mission control. The researchers found that crewmember mood and self-direction was reportedly greater in high autonomy, but mission control reported greater levels of anxiety and confusion in high autonomy. Of note, there were 4 Russian men and 2 European men in the simulation. The Russian crew members' task performance scores increased during high autonomy, whereas the European crew members' task performance scores decreased. Due to the small sample size and observed cultural differences, researchers could not isolate the effects of autonomy conditions and called for more research in long distance space exploration contexts with larger sample sizes [6].

Extant research has also investigated the effects of communication delays on individual and team behavior and performance in the ISS [7]. The study implemented a 50 s one-way communication delay between mission control and the 3-member ISS crew for parts of a 166-day mission. Crew well-being and communication quality significantly decreased in communication delay conditions compared with no communication delay controls. Kinz and colleagues [7] also found evidence to suggest that communication delays negatively impacted task efficiency, team/task coordination, and crew member stress levels. Taken together, extant research on autonomy and communication delay research suggests that individuals and teams are affected by some of the key features of space missions such as autonomy and communication delay [8].

Although the field's understanding of individuals and teams in space-related environments is increasing [9,10], less is known about how crews perform as a team on multiple dimensions. This research makes an important contribution by researching the capability of teams in isolated and confined settings to perform across multiple performance domains over time.

McGrath's task circumplex model [11] was used to delineate four aspects of team performance. High performing teams can: generate, choose, negotiate, and execute. Generate performance describes the degree to which a team can create original ideas and plans. Choose performance describes the degree to which a team can pool information to reach a decision that is correct from among a set of alternatives. Negotiate performance describes the degree to which a team can resolve conflicts among members. Finally, execute performance is the degree to which the team can complete behavioral tasks requiring the coordination of individual psychomotor tasks. The task battery assessed performance on the first three task domains, and NASA's MMSEV task was used to examine psychomotor performance [12].

Consider a scenario in which a space crew must quickly respond to an emergency situation without the required tools or parts using only supplies on board. This scenario would require the crew to come up with a novel or unknown solution. Generate tasks require crew member to engage in divergent thinking, moving away from known solutions to discover new ones.

Uncertainty and autonomy will also require crews to utilize their expertise to solve problems for which there is a correct solution. In the event of an emergency situation, a significant time delay will mean crew members may need to choose a solution before there is time to communicate the problem to mission control and receive the correct response. Crew members need to be able to effectively share knowledge and expertise with one another and be able to come to a group decision. Hence, crews will need to effectively perform choose tasks.

Another type of task that long distance space exploration crews will face are negotiate tasks. Negotiate tasks require the crew to resolve conflict. Astronaut Dave Williams describes a situation on the ISS where the crew debated how to orchestrate a spacewalk to repair a damaged tile. The crew had to determine the steps, roles, and timing, and each astronaut had definite ideas about how this should proceed [10]. There were multiple correct ways to do the spacewalk, but how the crew interacted to reach a decision was critical for the task.

Long distance crews will also have to perform well on execute tasks. Execute tasks require team members to coordinate psychomotor behaviors. Examples of these tasks are prevalent in space, for example, conducting EVAs, repairs, or remote operations of mechanical arms. Prior work on team performance in space crews has focused on execute performance.

Performance of four, 4-member analog space crews in NASA's HERA was tracked to understand how team performance is affected by time in isolation and confinement in a small group living scenario mimicking those of future long-distance missions.

2. Method

2.1. HERA 30-day study

NASA's HERA [13] resides at Johnson Space Center in Houston, Texas. In HERA, researchers study behavioral health, human factors, exploration medical capabilities, and communication and autonomy issues. HERA mimics the space mission environment in that crew members perform daily tasks, follow a minute-by-minute itinerary, exercise, and consume space food. Because HERA members cannot leave the analog and defer to a mission control, they also experience the isolation, confinement, and communication delays characteristic of space exploration missions. HERA participants are carefully selected based on similar characteristics required of astronaut personnel selection, such as age, height, physical and mental health, and advanced technical skills. During their time in isolation, participants are closely monitored to ensure physical and mental health and safety [13].

Data was collected from 4, 4-person crews, each participating in a 30-day mission in the research analog. The analog itself is an 80 m^3, three-story structure specifically designed to put participants into isolation, confinement, and remote conditions [13]. Crew members are confined to the inside of the capsule during the entirety of the mission. Level 3 consists of a small bunk for each crew member, in close proximity to the other 3 crew members. The entire crew shares a single “hygiene module”. There are two communal spaces in which activities such as experiments or exercise take place. Figs. 1–3 provide images of the HERA habitat. Participants who go into the analog are selected based on similarity in background to actual astronaut candidates (e.g., advanced degree in STEM field, experience leading missions in extreme environments on Earth, military flight experience).

The crew members in the study underwent 30-day missions in which the goal was to journey to the asteroid “Geographos” and collect rock samples before returning home. In order to reach the surface of the asteroid to collect rock samples, crew members participate in an EVA in which members take a MMSEV to the surface of the asteroid. Crew members spend mission days 1–15 in the outbound phase in which they simulate the trip to the asteroid. Crew members rendezvous with the asteroid on mission day 16 and spend the next few days conducting operations on the asteroid. Crew members then leave the asteroid on mission day 19 and are in the return phase of the mission from mission

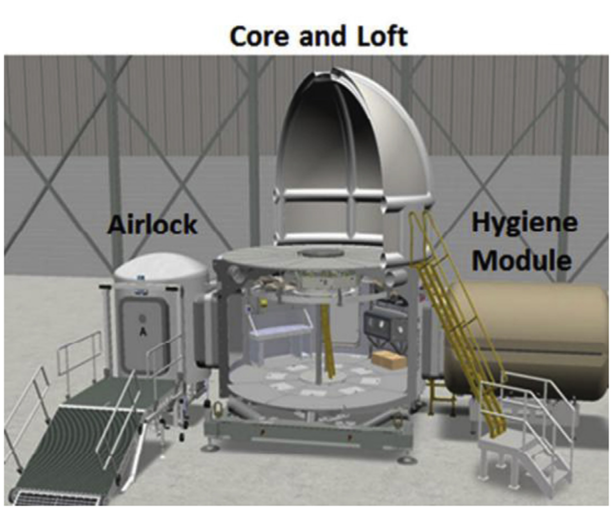

Fig. 1. a. Schematic representation of the HERA [13]. b. Schematic representation of the HERA [13].

Fig. 1. (*continued*)

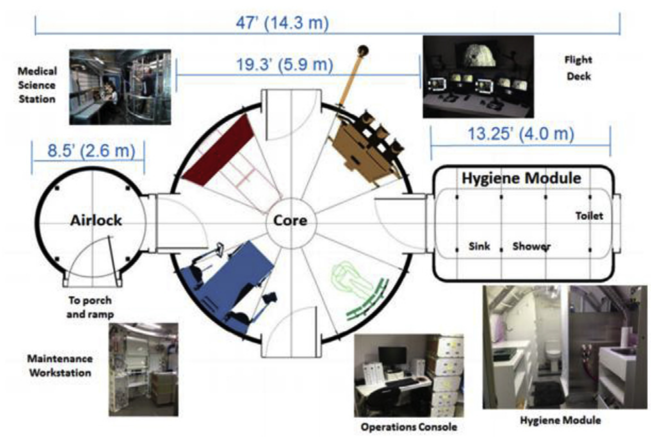

Fig. 2. HERA first level [13].

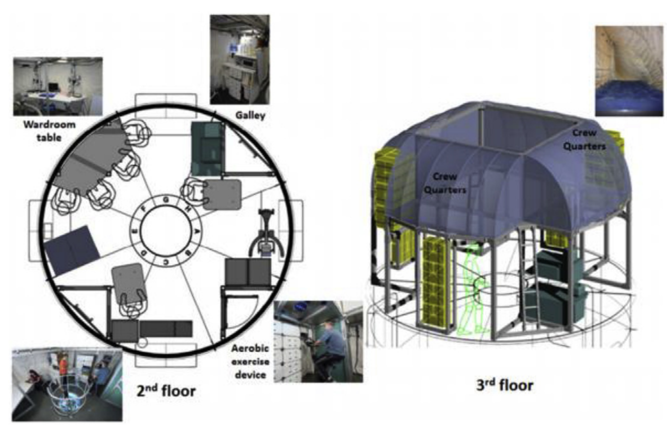

Fig. 3. HERA second and third levels [13].

days 19–30, returning to earth on day 30.

During the journey, researchers implemented a communication delay from mission day 13 to mission day 21. Specifically, there was a 30 s communication delay on mission days 13 and 21; a 1-min communication delay on mission days 14 and 20; a 3-min communication delay on mission days 15 and 19; and finally, there were three days of 5-min communication delay on mission days 16, 17, and 18. Crew members also underwent 24 h of sleep deprivation on mission day 24.

Table 1 presents details on each task type including assessment days in mission and scoring information. Participants completed the team task battery on days 11, 16, and 30. Day 11 was chosen as the start day to allow time for the crew to acclimate to the mission and to assess

Table 1
Tasks used to assess crew performance.

General Task Type	Specific Task in HERA	Assessed On MD:	Scoring Information
Generate: Creativity Task	Brainstorming	11, 16, 30	Scored using 3 dimensions - Fluency: the total number of non-repeating uses generated - Flexibility: the number of types of ideas generated - Novelty: the output dominance score (how often an item was listed relative to all items generated by all crews)
Choose: Intellective Task	Survival Analysis Task	11, 16, 30	Scored by computing the difference between the crew ranking of the item and the experts' rankings for each scenario
Negotiate: Cognitive Conflict Task	Ethical Dilemma	11, 16, 30	Scored using 2 dimensions: - Number of ideas: the total number of ideas and perspectives shared during the discussion - Discussion quality: SMEs evaluated each crew's discussion based on whether the scenario question was fully addressed, participant engagement in discourse, and consideration of various ideas
Execute: Psychomotor Task	MMSEV	1, 2, 3, 4, 7, 8, 9, 10, 11, 15, 16, 17, 18, 21, 22, 23, 24, 28	Scored by calculating the number of objectives completed out of the total assigned for each task.

performance prior to the initiation of communication delay. Day 16 was selected to assess performance during communication delay. Day 30 was chosen to assess crew performance after as long a period as was possible in this analog. Further, researchers intentionally avoided administering the battery in proximity to a 24-h sleep deprivation day. MMSEV data was collected on 18 of 30 days, with as many as 4 days between measures. There were slightly more assessments made in the first half of the mission, 10, than in the last half, 8. Table 1 presents the exact days of administration.

2.2. Tasks assessing team performance

In order to assess crew performance on the generate, choose, and negotiate dimensions, a battery was developed consisting of three parallel versions of a task assessing each dimension. The tasks were comparable to those used to assess collective intelligence along the same three dimensions, albeit in other contexts [14]. NASA included the MMSEV task, ideal for assessing execute performance as part of the standard mission procedure.

Creativity (generate) was defined as the production of ideas that are both novel and useful [15]. Creativity was assessed using an AUT approach [16]. The instructions were to generate as many uses for a specified object. Three parallel versions of the task were created by varying the object named in the task (i.e., brick, paperclip, rubber band). Crew members were instructed to spend 5 min individually generating alternative uses, and then came together as a crew to discuss additional uses. Tasks were completed utilizing Qualtrics, and it was through Qualtrics that crews received instructions and timing guidelines, and typed responses from tablets.

Creativity was scored using three dimensions: fluency, flexibility, and novelty, identified in past work on small group creativity [17–19]. Fluency was the total number of non-repeating ideas generated by the crew. Flexibility was the number of types of ideas generated for each crew. For example, a car stop and a door wedge counted as separate items in terms of fluency but represent one type of idea (using the item as a wedge). The number of categories for each item was determined by three raters categorizing all the items generated by the crews. If two of the three raters agreed on the category, the item was classified as that category [17]. If two raters did not agree, the item was discussed and a consensus reached. Finally, novelty [18,19] was determined by calculating an output dominance score - how often an item was listed relative to all items generated by all crews. Each item received a score and each crew's items were averaged to represent their novelty score.

The fluency, flexibility, and novelty scores were then standardized within dimension using the distribution of 4 crews each assessed at 3 times. The dimensions were strongly correlated: fluency and flexibility correlated 0.55, fluency and novelty 0.90, and flexibility and novelty 0.52. The standardized scores on fluency, flexibility, and novelty were averaged and used in subsequent descriptive analysis of team creativity.

Problem solving (choose) was defined as a team's ability to select a demonstrably correct answer [11,20]. Problem solving was assessed using intellective tasks that present participants with survival scenarios such as the classic NASA moon survival task [21]. Each participant was provided with an explanation of the situation and a list of 15 available items, which they were to rank in order of their importance 1 (most important) to 15 (least important) to crew survival. Crews completed the tasks using instructions presented in Qualtrics. The instructions were to spend 10 min independently reviewing the scenario and ranking the items. Next, the crew members were instructed to spend 15 min discussing their rankings and arrive at a final crew ranking that represents their best assessment of the importance of the items. Three parallel versions of the task were created by varying the survival scenario and objects ranked by the crew. The first scenario was the NASA moon survival task with validation provided by NASA experts. The second scenario was a desert survival task with validation provided by the Chief of the Desert Branch, Tropic Information Center of the Air Force University at Maxwell Air Force Base. The third scenario was a winter survival task validated by a US Army survival trainer.

Problem solving was scored by computing the difference between the crew ranking of the item and the respective expert ranking. Problem solving scores were computed using the absolute value of the deviation of each item's crew assigned rank from the expert ranking, and then summing the deviations. Because greater deviation reflects lower performance, researchers reverse scored this measure by converting scores to negative numbers, making zero the best possible score and −112 the worst possible score. For example, imagine if a magnetic compass and a flashlight were ranked 5 and 13, respectively, by the crew. However, subject matter experts ranked the magnetic compass and the flashlight as 13 and 5, respectively. In the crew's case, the two items would both be scored as 8, giving the crew a score of −16. These scores were then standardized across crews and time points, and the resulting standard score was used in subsequent descriptive analysis.

Ethical decision making (negotiate) was categorized as a subjective task influenced by values and attitudes rather than facts [11,21]. Ethical decision making was assessed using three moral dilemmas. The first dilemma was a runaway trolley that was heading towards a group of people, and the crew has the ability to divert the trolley, but doing so would result in the certain death of one person [22]. The second dilemma involved the trade-off of building a vital pipeline, but doing so would come at the cost of human life [23]. In the third dilemma, a ship has struck an iceberg and there were not enough life rafts for all the passengers; the crew had to decide who would be given access to the life rafts and whether they should be held personally responsible for lives lost due to the decision [24]. All three tasks were completed via

Qualtrics. The crew was given 20 min to read and discuss the dilemma and 5 min to type their decision into the Qualtrics survey.

Ethical decision making was scored by three coders who reviewed video recordings of the crew discussion. Coders calculated the total number of ideas and perspectives shared during the discussion. Additionally, coders evaluated each crew's discussion quality based on the scenario question being fully addressed, participant engagement in discourse, and consideration of alternative ideas. An overall discussion quality score of 1 (poor) to 5 (excellent) was assigned to each crew on each task. The number of ideas generated and overall discussion quality were highly correlated (r = 0.69), and scores were standardized and averaged to comprise one ethical decision making score per crew at each time point. The resulting aggregate was used in subsequent descriptive analysis.

The MMSEV/EVA task (execute) was an objective performance task in which the four crew members participated in a virtual reality space simulation. During this task, the crew worked together to complete objectives at specific locations on an asteroid. The pilot and co-pilot remained on the MMSEV for the duration of this task. They piloted the MMSEV between the ship and the asteroid, as well as direct the other two crew members to each objective by using a painting-laser. The other two crew members departed the MMSEV upon arrival at the asteroid. They proceeded to the coordinates of each assigned objective to collect various types of rock samples or set up sensor relays. Data that showed how the team attempted to complete these objectives were hand recorded by NASA operations personnel during this task.

Three undergraduate research assistants coded the MMSEV performance data. Each coder compared the data to the teams' assigned objectives for a given mission day and coded each objective as "complete," "partial completion," or "no evidence for completion." The tasks were divided between the three coders so as to have each task coded twice. The coders agreed on the majority of data; of 1098 total codings, coders agreed on 902. Joint probability of agreement was calculated at 82.15%. All discrepancies between codings were resolved using consensus between the three undergraduates and a graduate student task expert.

3. Results

Table 2 presents the descriptive statistics on crew performance dimensions for the four crews. Examining generate performance, there was wide variation in creativity, with the high score of 45 ideas more than double the low score of 22 ideas. Similar variation was observed with the number of kinds of ideas, and in novelty as well, though the scaling of the novelty ratings makes this difficult to see in the descriptives. There was wide variation in choose performance; the best choose performance was a deviation of 24 points from the expert model across 15 items, whereas the worst observation was 66 points. This is a meaningful performance difference, ranging from being inaccurate on each item's survival utility by 1–2 points (best team score) versus innacturate by 4–5 points (worst team score). Execute performance showed similarly wide variation, ranging from 47% of objectives completed to 98% of objectives completed.

Table 2
Descriptive statistics for HERA Campaign 3 crew performance (*N* = 12 observations).

Variable	Mean	SD	Minimum	Maximum
1. Generate				
Fluency	36.08	8.05	22	45
Flexibility	17.83	3.97	13	25
Novelty	0.98	0.01	0.97	0.99
2. Choose	−46.83	11.49	−66	−24
3. Negotiate				
Number of ideas	8.17	3.35	5	16
Discussion quality	3.71	0.96	2	5
4. Execute	0.85	0.15	0.47	0.98

Note. Reported means and standard deviations are calculated from the raw score values for each dimension.

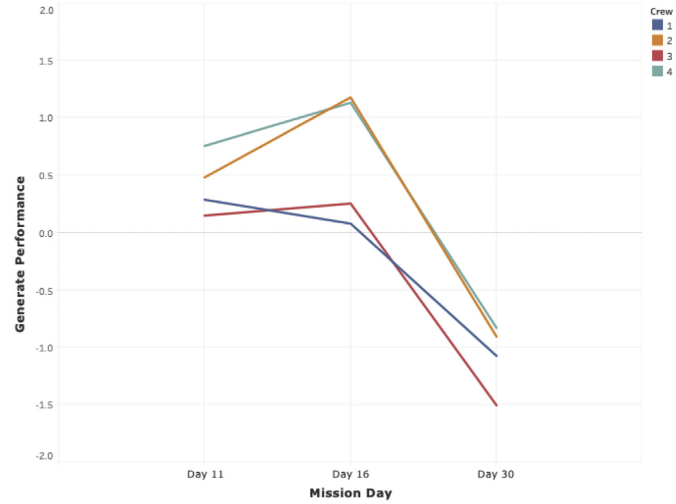

Fig. 4. Generate performance over time.

For interpretability across different measurement scales, the raw scores were converted to z-scores using the distributions of scores for the four crews at the three time points. This process was done for each of the 4 performance dimensions (generate, choose, negotiate, and execute). The plots presented for each dimension include the z-scores.

Fig. 4 displays the crews' generate performance. Fig. 4 depicts a slight increase from mission day 11 to mission day 16 for 3 out of 4 of the crews. However, all 4 crews show a steep decrease in performance from mission day 16 to mission day 30.

Fig. 5 displays crews' choose performance. Fig. 5 depicts the problem-solving scores of each of the crews at each of the three time points. Fig. 6 shows all crews decreased in problem solving performance from mission day 11 to mission day 16. From mission day 16 to mission day 30, there is no distinct pattern of performance across crews. Two crews increased in problem solving ability while two crews decreased in problem solving ability.

Fig. 6 displays crews' negotiate performance. Fig. 6 shows the ethical decision-making scores of each of the crews at each of the three time points. Ethical decision making shows a relatively stable performance trend. Crew 3, however, exhibited an increase in ethical decision-making performance from day 16 to day 30.

Fig. 7 displays crews' execute performance. Fig. 7 shows the degree

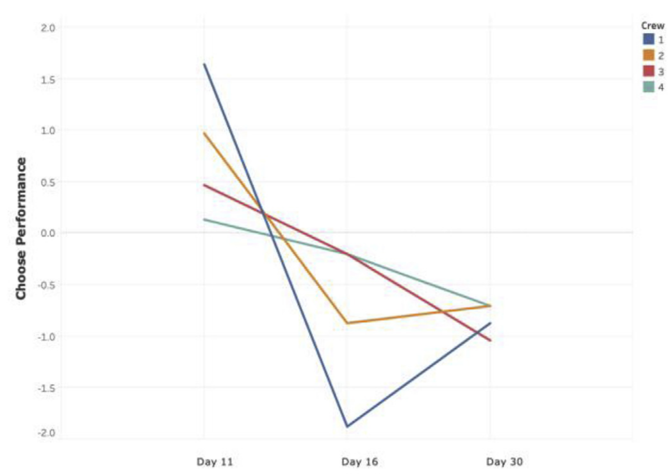

Fig. 5. Choose performance over time.

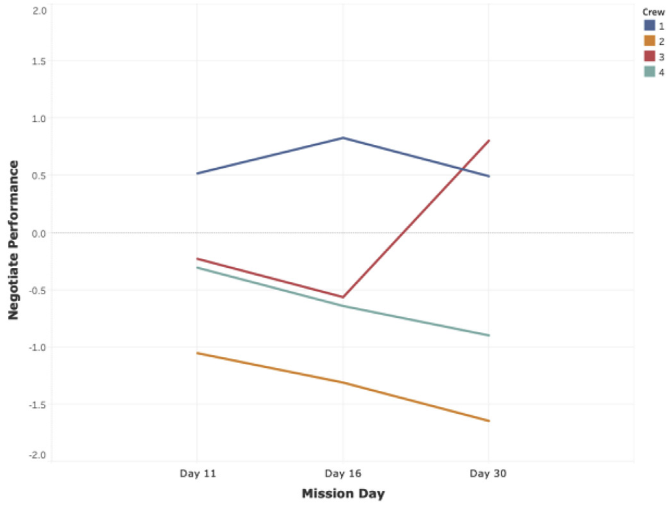

Fig. 6. Negotiate performance over time.

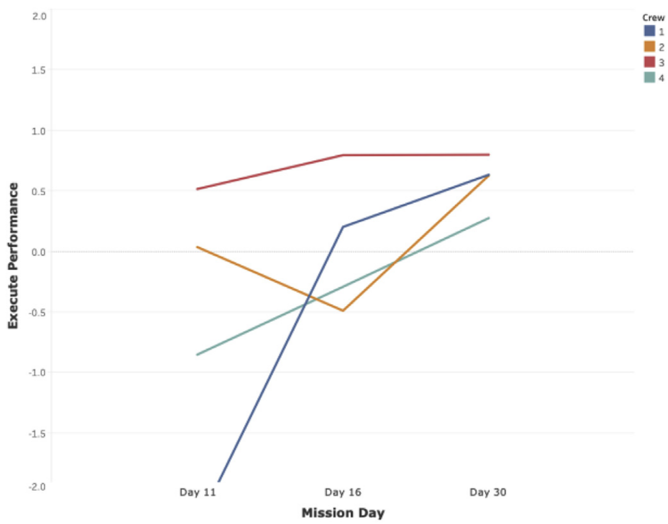

Fig. 7. Execute performance over time.

to which each crew met their total number of assigned objectives on a given mission day, shown here as a z-sscore. Crew 1 had the largest change in performance; they achieved very low performance at the beginning of the mission but had the second highest by mid-mission. Crews 2 and 4 started in the below average to average range, and increased their performance steadily throughout the mission. Crew 3 had the highest performance of all 4 crews, starting with the best performance and maintaining this trend throughout the mission. It is of note that ceiling effects may have distorted the true performance of Crew 3, as they achieved perfect or near perfect scores on the MMSEV task for the majority of the mission.

4. Discussion

Crew members of future long-distance space exploration missions are poised to face more novel environments and challenges than ever before [25]. Such crews will have unprecedented levels of autonomy. The ISS model where the: "Crew is [Mission Control's] eyes… Ground is the brain," (p. 607) [26] does not translate well to long distance missions. Future crews need to be both the eyes *and* the brain.

Crew members will have to work well with each other in both routine and novel situations on a wide spectrum of task types, from creativity work to highly complex intellective issues to ethical dilemmas to psychomotor challenges. This study examined the effects of isolation and confinement on four essential dimensions of team performance over time. The results demonstrate that measuring multiple performance dimensions is a useful practice and should be strongly considered for analog research and long duration mission monitoring alike. Certain dimensions of performance may not vary at the same rate or in the same direction. Thus, each of the four team performance dimensions should be assessed.

The results have important implications for future research and the management of future space crews. The study showed differential effects for each of the four performance dimensions over time. Although a focus on relations between crew members is important to ICE and space contexts, the risk of mission failure comes specifically from the crew not performing at adequate levels.

Whereas execute and negotiate performance were stable or increased over time, generate and choose performance declined over time. Execute tasks may have increased over time as the crew learned the task and how to work together. Execute tasks are largely behavioral, requiring members to coordinate joint actions. In contrast, generate and choose tasks are conceptual. These task dimensions may have shown decrements as members became less communicative in isolation [27]. Conceptual tasks require perspective-taking, empathy, reasoning, and elaboration of information to teammates, all of which may suffer when individuals begin to withdraw in isolation. Negotiate performance is the most process-intensive, and it showed the greatest stability. It may be the case that crews differ from each other in the quality of their interaction processes, and that these abilities remain relatively stable over time.

The current findings underpin the importance of considering crew performance on conceptual tasks in future analog studies. The variations and differences of specific dimensions of team performance over time suggest that the dynamics of team performance are critical.

These results indicate the importance of broader conceptualization and measurement of team performance. Additionally, while team performance is made up of individual and relational processes, the crew will succeed or fail as a group. Synergies and process losses render team performance more than the sum of individual crew members' performance levels, and so tracking team level indicators is essential.

In fact, it is these conceptual dimensions of performance that largely distinguish the current era of long duration space missions like the ISS from the next era of long-distance missions to Mars and beyond. The crew autonomy created by physical distance will require the crew to perform exceptionally well across all four dimensions for the entirety of the mission. When considering how to best apply results from analog studies, one must consider the differences between the analog mission and the missions which will actually be undertaken by astronauts in the future. While range restriction of longitudinal performance data might impact the predictive utility for longer missions, it is likely that the performance decrements seen in our study would increase over time. In other words, if performance on these tasks was easy to maintain over time, we would expect them to be better for a 30-day mission. When making predictions about their performance on longer missions, it is unlikely crews would improve performance on these dimensions in the face of the extended hardships of space travel. If a crew cannot maintain performance for 30 days, how can we expect them to maintain performance for 3 years?

The observations of this sample can be used to build on and extend conclusions based upon other studies of crews in space analogs. For example, one such study of the Mars 105 mission found the crew felt a decrease in positive emotions throughout the mission, and that certain coping strategies might be better for preventing symptoms of depression [28]. As these coping strategies could be taught to astronauts, it may also be enlightening to conduct an examination of whether team interventions might build resiliency against the performance decrements observed in our analysis. Ultimately, it is our hope that these results will add to a thriving research stream which will shed light on how to compose crews who are best equipped to handle the stressors of long duration space travel.

5. Conclusion

In conclusion, results from this study tracking teamwork over time reveal that it is essential to monitor crew performance on conceptual team tasks and to use these as criteria in studies developing recommendations and countermeasures for long distance missions. Whereas many aspects of astronaut selection and training focus on highly behavioral coordinative tasks that fall in the execute domain, this sample demonstrates decrements in the more interpersonally and knowledge-intensive domains of generate and choose performance. These dimensions are likely to be critical as astronauts on future space missions becomes more reliant on one another to solve the inevitable and unanticipated challenges and opportunities of exploring deep space.

Acknowledgements

This research is based upon work supported by NASA under awards NNX15AM32G (Project CREWS), 80NSSC18K0276 (Project CREST), and 80NSSC18K0221 (Project SCALE). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Aeronautics and Space Administration. We thank Gabriel Plummer, Ashley Niler, and Alexa Harris for their assistance with this research.

References

[1] Jet Propulsion Laboratory (n.d.), Mission timeline: surface operations, Retreived from https://mars.jpl.nasa.gov/mer/mission/tl_surface_nav.html.
[2] L.L. Schmidt, K. Keeton, K. Slack, L.B. Leveton, C. Shea, Risk of performance errors due to poor team cohesion and performance, inadequate selection/team composition, inadequate training, and poor psychosocial adaptation, in: J. McMurphee, J. Charles (Eds.), Human HHealth and Performance Risks of Space Exploration Missions (NASA: SSP-2009-3405), National Aeronautics and Space Administration, Washington, DC, 2009.
[3] S.W. Kozlowski, D.R. Ilgen, Enhancing the effectiveness of work groups and teams, Psychol. Sci. Publ. Interest 7 (3) (2006) 77–124.
[4] K. Dismukes, B.A. Berman, L.D. Loukopoulos, The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents, Ashgate Publishing, Ltd, 2007.
[5] S.T. Bell, S.G. Brown, D.R. Abben, N.B. Outland, Team composition issues for future space exploration: a review and directions for future research, Aero. Med. Hum. Perform. 86 (6) (2015) 548–556.
[6] N. Kanas, M. Harris, T. Neylan, J. Boyd, D.S. Weiss, C. Cook, S. Saylor, High versus low crewmember autonomy during a 105-day Mars simulation mission, Acta Astronaut. 69 (2011) 240–244.
[7] N.M. Kinz, C. Chou, W.B. Vessey, L.B. Leveton, L.A. Palinkas, Impact of communication delays to and from the International Space Station on self-reported individual and team behavior and performance: a mixed-methods study, Acta Astronaut. 129 (2016) 193–200.
[8] S.J. Golden, C. Chang, S.W.J. Kozlowski, Teams in isolated, confined, and extreme (ICE) environments: review and integration, J. Organ. Behav. (2017) 1–15.
[9] S.T. Bell, S.G. Brown, T. Mitchell, What we know about team dynamics for long-distance space exploration missions: a systematic review of analog research, Front. Psychol. 10 (2019) 811.
[10] M. Robson, D. Williams, L. DeChurch, 'It's like a hall of mirrors, A Spacecraft, Some Personalities Work Better than Others (L. Lynch, Interviewer) [Audio File], 2019 Retrieved from https://www.cbc.ca/radio/thecurrent/the-current-for-february-22-2019-1.5029360/friday-february-22-2019-full-transcript-1.5030524.
[11] J.E. McGrath, Groups: Interaction and Performance, Prentice Hall, Englewood Cliffs, NJ, 1984.
[12] M. Reagan, B. Janoiko, J. Johnson, S. Chappell, A. Abercromby, NASA's analog missions: driving exploration through innovative testing, AIAA SPACE 2012 Conference & Exposition, American Institute of Aeronautics and Astronautics, Pasadena, California, 2012, , https://doi.org/10.2514/6.2012-5238.
[13] R.L. Cromwell, J.S. Neigut, Human Research Program Human Exploration Research Analog (HERA) Experiment Information Package, National Aeronautics and Space Administration, Washington, DC, 2014.
[14] A.W. Woolley, C.F. Chabris, A. Pentland, N. Hashmi, T.W. Malone, Evidence for a collective intelligence factor in the performance of human groups, Science 330 (6004) (2010) 686–688.
[15] T.M. Amabile, The social psychology of creativity: a componential conceptualization, J. Personal. Soc. Psychol. 45 (2) (1983) 357–376.
[16] J.P. Guilford, The Nature of Human Intelligence, McGrawHill, New York, 1967.
[17] C.T. Tadmor, A.D. Galinsky, W.W. Maddux, Getting the most out of living abroad: biculturalism and integrative complexity as key drivers of creative and professional success, J. Personal. Soc. Psychol. 103 (3) (2012) 520.
[18] J.C. Kaufman, R.J. Sternberg (Eds.), The Cambridge Handbook of Creativity, Cambridge University Press, 2010.
[19] J.G. Lu, M. Akinola, M.F. Mason, "Switching On" creativity: task switching can increase creativity by reducing cognitive fixation, Organ. Behav. Hum. Decis. Process. 139 (2017) 63–75.
[20] S.G. Straus, Testing a typology of tasks: an empirical validation of McGrath's (1984) group task circumplex, Small Group Res. 30 (2) (1999) 166–187.
[21] J. Hall, W.H. Watson, The effects of a normative intervention on group decision-making performance, Hum. Relat. 23 (4) (1970) 299–317.
[22] P. Foot, Theories of Ethics, Oxford University Press, 1967.
[23] V. Grassian, Moral Reasoning: Ethical Theory and Some Contemporary Moral Problems, Prentice Hall, Englewood Cliffs, NJ, 1981.
[24] W.H. Patterson Jr., Robert A. Heinlein: in Dialogue with His Century: 1948-1988 the Man Who Learned Better, Macmillan, 2014.
[25] E. Salas, S.I. Tannenbaum, S.W. Kozlowski, C.A. Miller, J.E. Mathieu, W.B. Vessey, Teams in space exploration: a new frontier for the science of team effectiveness, Curr. Dir. Psychol. Sci. 24 (3) (2015) 200–207.
[26] J. Mesmer-Magnus, D.R. Carter, R. Asencio, L.A. DeChurch, Space exploration illuminates the next frontier for teams research, Group Organ. Manag. 41 (5) (2016) 595–628.
[27] L.A. Palinkas, J.C. Johnson, J.S. Boster, Social support and depressed mood in isolated and confined environments, Acta Astronaut. 54 (9) (2004) 639–647.
[28] M. Nicolas, G.M. Sandal, K. Weiss, A. Yusupova, Mars-105 study: time-courses and relationships between coping, defense mechanisms, emotions and depression, J. Environ. Psychol. 35 (2013) 52–58.