72nd International Astronautical Congress (IAC), Dubai, United Arab Emirates, 25-29 October 2021.

IAC-21-A1.1.4

Re-pairing to Repair: A Countermeasure that Enhances Crew Relations in Deep Space

Brennan Antone[a]*, Alina Lungeanu[b], Leslie A. DeChurch[b], Suzanne T. Bell[c], Noshir S. Contractor[abd]

[a] *Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA,* brennanantone@u.northwestern.edu; nosh@northwestern.edu
[b] *Department of Communication Studies, Northwestern University, 2040 Campus Drive, Evanston, Illinois, 60201, USA,* alina.lungeanu@northwestern.edu; dechurch@northwestern.edu
[c] *Behavioral Health and Performance Laboratory, NASA, 2101 E NASA Pkwy, Houston, TX 77058, USA* suzanne.t.bell@nasa.gov
[d] *Department of Management and Organizations, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, USA*

* Corresponding Author

Abstract

Supporting effective social relationships is critical for teams - especially the crews of space exploration missions, who must collaborate on tasks autonomously under extreme conditions that have been shown to degrade social relationships. We develop an agent-based computational model (ABM) that simulates social relationships in crews as they complete their work, as well as the evolution of these social relationships during long-duration space exploration (LDSE). We calibrate the model using data gathered from crew completing missions in LDSE analogs, on Earth, over a period of 30- to 45-days. This empirical calibration allows the model to predict the formation of social relationships based on team composition as well as task scheduling and LDSE characteristics. We demonstrate how empirically driven ABMs enable the recommendation of countermeasures that can support teams, in this case to promote positive and dissuade negative social relationships within teams. We consider a "*re-pairing*" countermeasure, in which we reassign which pairs of crew members are assigned to collaborate based on simulation results from the ABM. We test the effectiveness of these ABM-recommended countermeasures among four member crews completing 45-day missions in the HERA LDSE analog. Crews are assigned to work on tasks in either "recommended" or "worst" pairs determined by the ABM. Use of "recommended" or "worst" pairs was alternated each quarter of the mission, as part of an "A-B-A-B" block experimental design within each crew. During the "recommended" pairings, the crew members were more likely to report positive social relationships and were less likely to report that working together was damaging to their relationship. These findings demonstrate the viability of using computational models to recommend team countermeasures for deep space exploration.

Keywords: social networks, agent-based model, computational model, countermeasures, interventions, task schedule

Acronyms/Abbreviations

Agent-Based Model (ABM)
Crew Recommender for Effective Work in Space ABM (CREWS)
Human Exploration Research Analog (HERA)
Isolated, Confined, and Controlled Environments (ICC)
Long-Duration Space Exploration (LDSE)
National Aeronautics and Space Administration (NASA)
NASA Task Load Index scale (NASA-TLX)
Team Task Analysis scale (TTA)

1. Introduction

Human exploration of deep space will require unprecedented levels of crew autonomy. Under increased communication delay that comes with more distant explorations, crew members will have to function more independently from mission control. At the same time, these explorers will live and work for extended periods of time in a small space and apart from loved ones, all while facing extreme physical challenges associated with gravity and radiation. Past research on extreme teams has already demonstrated that these conditions strain interpersonal relationships necessary to autonomous team performance. The purpose of our research is to develop and test a computational model capable of recommending countermeasures that support crew social relationships in deep space.

Computational modeling is a technique designed to mimic complex systems. It provides researchers the ability to understand, predict, and plan interventions into real-world systems. An agent-based model (ABM) is one such type of model. ABMs break complex systems (*e.g.,* teams) down into parts called agents (*e.g.*, team members, tasks). By simulating local interactions between agents, they model both local and emergent global patterns in systems. Social scientists create ABMs that explain a variety of processes occurring within teams or organizations by representing sequences of interactions between individuals. Researchers are able to incorporate empirical data into model design, calibration, and testing in order to help establish a model's validity.

We developed and validated the CREWS computational model on data from 4-person crews living in NASA's Human Exploration Research Analog (HERA) for either 30 or 45 days [1]. The model simulates interpersonal relationships between crew members based on their personal characteristics gathered before the mission, along with their mission timeline and their task schedule. CREWS anticipates the different ways that crew relationships might change each mission day. For instance, individuals who are high in self-monitoring are less prone to developing negative relationships, but high workload days promote negative relationships. By performing *in silico* simulations of potential countermeasures, the model evaluates which options would be most effective at mitigating risks related to changing crew relationships.

We tested the computational model's effectiveness at recommending one type of countermeasure: *crew re-pairing.* This countermeasure consists of changing which sets of crew members are assigned to work as a pair on the most interdependent tasks. Given several potential task assignments, our model predicts how these assignments impact crew relationships. These predictions are then used to adjust the schedule so that it incorporates personal, relational, situational, and operational demands. By using a block experimental design, we tested in Hera whether the crew re-pairing countermeasure decisions recommended by the ABM were effective.

2. Methods

2.1 Long-Duration Space Exploration Analog

To study crews in environments that mimic the extreme characteristics of work in space, LDSE analogs are operated on Earth. HERA is a LDSE analog operated by NASA that places four-member crews in a small habitat to live and work together over 30 to 45 days, while contending with social isolation, communication delays, workload, slam shifts, and sleep deprivation designed to mimic space exploration. Figure 1 depicts the HERA module.

Fig. 1. Exterior of the HERA module in which crew members work and reside to simulate LDSE. *Photo credit: NASA*

2.2 Sample

We used two samples of participants. The first sample was used in the development and calibration of our agent-based model (ABM). This sample (n = 32) consisted of the participants in eight four-member crews completing missions in HERA. Four of these crews completed 30-day missions in HERA as part of Campaign 3 of the analog. The remaining four crews completed 45-day missions in HERA as part of Campaign 4. The second sample was used in our experimental research design testing countermeasures implemented based on our ABM. This sample (n = 16) consisted of the participants in four four-members crews completing missions in HERA. These crews completed 45-day missions in HERA as part of Campaign 5.

2.3 CREWS Agent-Based Model (ABM)

The CREWS (Crew Recommender for Effective Work in Space) ABM was designed to predict social relationships between crew members during a long-duration space exploration mission. As a computational model, it simulates dynamic changes in social relationships over time. The model focuses on four types of social relationships as criterion variables. Task Affect ("With whom do you enjoy working?") represents positive working relationships, while Task

Hindrance ("Who makes tasks difficult to complete?) represents negative relationships. Additionally, the model examines perceptions of Leadership ("To whom do you provide leadership?") and Followership ("Who do you rely on for leadership?"). These four network ties were shown to be connected to team performance in LDSE in previous work [2]. These social relationships are modeled as networks of directed, binary ties between crew members. Measurement of these relationships is described in Appendix A.

To predict these four types of relationships, the ABM integrates different types of factors that have been shown to influence social relationships in LDSE-analogs specifically, and in teams generally. First, factors that are unique to Isolated, Confined and Controlled (ICC) environments that occurred in HERA were modeled: communication delays, sleep deprivation, high workload days, and total time spent in isolation. Next, the specifics of each individual task scheduled on each day were considered: workload of tasks, the level of interdependence when team members worked on the task together, situational strength when performing the task, and the tasks duration. Additionally, individual differences of crew members (i.e., Five Factor personality traits, values, coping styles, psychological collectivism, and self-monitoring) were included. Both main effects of personality, as well as similarity between the personality traits of multiple crew members, were considered. Finally, network effects - how the relationship between one pair of crew members affects the relationships between other crew members - were included in the model. Figure 2 summarizes how the ABM integrates the effects of multiple factors on the development of social relationships.

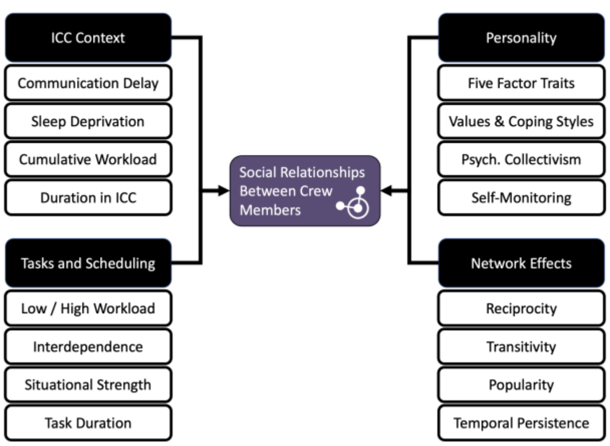

Fig. 2. Conceptual model of multiple factors integrated into CREWS ABM.

The ABM works by simulating shifts in social networks (task affect, hindrance, leadership, followership) as the crew completes different tasks throughout the course of a mission. Each time a subset of crew members work together on a task, as scheduled in HERA, the ABM updates all social ties between that subset of crew members. Ties in all four social networks are binary - they either do or do not exist. To update ties, the ABM determines whether a tie forms where it did not exist previously, or whether a tie that previously existed dissolves. Figure 3 depicts what the interface of the model looks like at a single point in time, taking the case of having the two mission specialists (blue) completing exercise tasks together.

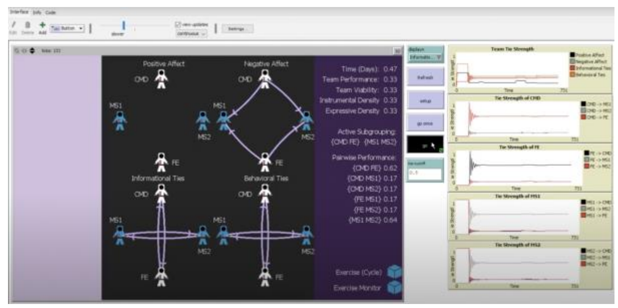
Fig. 3. Example of ABM interface, depicting the evolution of four types of network ties over time.

Whether a tie forms or dissolves at each step is stochastic, modelled as a logistic function of the predictor variables included in the model (ICC context, task attributes, individual differences such as personality traits, and network effects). Different predictor variables have different magnitudes of effects on ties in each of the four social networks. The magnitude of each variables' effect was determined by calibrating the model using empirical data gathered from the first eight crews in our sample (n = 32). More information on the design and development of the CREWS ABM, including the results of calibrating the model to data gathered from our first eight crews, is detailed in [1].

2.4 Task Pairing Countermeasures

A potential use of the CREWS ABM, particularly related to its ability to predict social relationships, is to conduct *in silico* simulations to observe what social relationships will form in different hypothetical scenarios. If our goal is to help crew members form positive relationships, and to help them avoid negative relationships, then the ABM is able to act as a decision support tool. For instance, when selecting team members (team composition) several potential teams could be input into the ABM. Then, the ABMs predictions about the social relationships that would form in each team could be used to select the team with the greatest potential for success. As another example, when scheduling tasks for crews to complete, the ABM allows us to compare several potential task schedules, and to identify what different effects the schedules may have on the social relationships between crew members.

To test the ability of the ABM to recommend countermeasures that support positive relationships and prevent negative relationships, we use it to make a simple scheduling decision. Specifically, within HERA, there are several two-person tasks that crew members split into pairs to complete them. This allows us to consider which pairs of crew members should be assigned to work together for two such tasks: the *Rover* task and the *Phobos Sampling* task.

In the *Rover* task, pairs of crew members work on their own small robotic vehicle as part of an educational outreach activity, as shown in Figure 4. The story behind the task is that the robotic vehicle was designed by a fictitious high school robotics team, and the pair of crew members must work to assemble, run tests on, and reconfigure the rover. The task is designed to be stressful and challenge communication skills, by incorporating confusing procedures and requiring the crew members to communicate externally with mission control. The Rover task is a 60-minute task performed on mission days 4, 6, 13, 19, 21, 25, 32, 33, 35, and 42 of the mission. In an analysis of the 43 task categories crews performed in HERA, we found that the Phobos Sampling task was in the 95th percentile in terms of workload (Nasa-TLX) and in the 77th percentile in terms of team interdependence (TTA). More detail on the comparison of task attributes can be found in Appendix C.

Fig. 4. Images from crew members completing the Rover task. *Photo credit: NASA*

The *Phobos Sampling* task involves the simulated geological analysis of samples from the Martian moon Phobos, as shown in Figure 5. Pairs of crew members worked on retrieving samples from storage, transferring samples to a glove box, analyzing samples, and returning them to storage. The task entails careful handling of samples, collection of simulated data, and operation of equipment alongside antiseptic and containment techniques. Phobos Sampling is a 60-minute task that was conducted exclusively in the second half of the mission (when the simulated mission has arrived at Mars), on days 32 and 38 of the mission. In an analysis of the 43 task categories that crews performed in HERA, we found that the Phobos Sampling task was in the 82nd percentile in terms of workload (Nasa-TLX) and in the 84th percentile in terms of team interdependence (TTA). More detail on the comparison of task attributes can be found in the Appendix C.

Fig. 5. Images from crew members completing the Phobos Sampling task. *Photo credit: NASA*

Both tasks represent a small minority of interactions occurring within HERA. The crews are living and working together full time for 45 days. Collectively, these two tasks take up only a total of 12 hours over the course of the mission. Thus, we expect the effect of assigning pairings to complete these tasks to have a relatively small effect on the crew's social relationships. However, the two tasks are among the few tasks in HERA in which assigned pairs complete a work task (rather than solo or four-member tasks), while also being high in workload and task interdependence, two attributes that confer them a salient effect on social relationships.

2.5 Recommending Task Pairings using the ABM

To test the ability of our ABM to recommend countermeasures, we used the ABM to simulate one decision: How should the four-member crew be split into two pairs when completing both the Rover and the Phobos Sampling tasks? There were three possible ways of assigning the crew members to pairs: (i) pairing the Commander with the Flight Engineer, and pairing Mission Specialist 1 with Mission Specialist 2, (ii) pairing the Commander with Mission Specialist 1, and pairing the Flight Engineer with Mission Specialist 2, or (iii) pairing the Commander with Mission Specialist 2, and pairing the Flight Engineer with Mission Specialist 1. The ABM was limited in its ability to select countermeasures by having only three options, and the need to have each crew member be paired with someone to work on the task.

Prior to the start of the HERA mission, data was gathered to allow our ABM to predict social relationships: crew member demographics, personality measures, and the scheduling of tasks in HERA. Three versions of the schedule were considered, each being identical except for the choice of which crew members were paired together on the Rover and the Phobos Sampling tasks. This allowed the ABM to simulate what social relationships were predicted to look like under all three schedules, prior to a crew's actual egress.

For each potential crew schedule, our model performed 200 simulations predicting each crew's social relationships over the mission. Because of the stochastic nature of our model, each of these 200 simulated crews had slightly different relationships. Within each crew, we then compared the relationships across the schedules with the three different task pairings (*e.g.* Crew 1's relationships when the Commander was paired with the Flight Engineer, vs. Crew 1's relationships when the Commander was paired with Mission Specialist 1, vs. Crew 1's relationships when the Commander was paired with Mission Specialist 2).

Observing 200 sets of simulated relationships allowed us to identify one of the three ways of pairing crew members for the Rover and Phobos Sampling tasks as the "recommended" pairing and identified one of them as the "worst" pairing. These pairings were selected through examining simulated task affect, hindrance, leadership, and followership networks at each day of the mission. The pairing options that maximized the predicted task affect relationships and minimized the predicted hindrance relationships were labeled as "recommended". "Worst" pairings were those who minimized task affect relationships and maximized hindrance.

2.6 Experimental Design

To test whether the ABM-recommended pairings were effective, we implemented pairings in 4 crews completing 45-day missions in HERA. We utilized a block experimental design to make within-crew comparisons between the "recommended" pairings and the "worst" pairings identified by our model.

The 45-day mission was split into four quarters, the mission was split into four quarters and crews were assigned to alternate between working in either their "recommended" or their "worst" pairings each quarter, following an "A-B-A-B" or "B-A-B-A" block experimental design. The Rover task was completed twice in the first quarter (days 1-11), three times in the second quarter (days 12-22), twice in the third quarter (days 23-34), and twice in the fourth quarter (days 35-45). The Phobos Sampling task was completed only once in the third quarter and once in the fourth quarter of the mission. The quarters that the members spent working in their "recommended" and "worst" pairings for each crew are depicted in Table 1.

Table 1. Timeline of using "recommended" and "worst" pairings for each crew in each quarter of the mission.

	Quarter 1: Days 1-11	Quarter 2: Days 12-22	Quarter 3: Days 23-34	Quarter 4: Days 35-45
Crew 1	**Best pairing** (CMD-MS1) (FE-MS2)	**Worst pairing** (CMD-FE) (MS1-MS2)	**Best pairing** (CMD-MS1) (FE-MS2)	**Worst pairing** (CMD-FE) (MS1-MS2)
Crew 2	**Worst pairing** (CMD-MS2) (FE-MS1)	**Best pairing** (CMD-MS1) (FE-MS2)	**Worst pairing** (CMD-MS2) (FE-MS1)	**Best pairing** (CMD-MS1) (FE-MS2)
Crew 3	**Best pairing** (CMD-FE) (MS1-MS2)	**Worst pairing** (CMD-MS1) (FE-MS2)	**Best pairing** (CMD-FE) (MS1-MS2)	**Worst pairing** (CMD-MS1) (FE-MS2)
Crew 4	**Worst pairing** (CMD-MS2) (FE-MS1)	**Best pairing** (CMD-FE) (MS1-MS2)	**Worst pairing** (CMD-MS2) (FE-MS1)	**Best pairing** (CMD-FE) (MS1-MS2)

We conducted a Countermeasures survey to assess the effect that the task pairings had upon the social relationships of crew members within 72 hours following each instance of the Rover or the Phobos Sampling task. The survey included two-item scale assessing Constructive Effects of the Rover task (α = 0.95, *e.g.* "Spending time with my partner on the Rover task was helpful to our relationship"), Destructive Effects of the Rover task (α = 0.89, e.g. "Spending time with my partner on the Rover task was damaging to our relationship"), Constructive Effects of the Phobos Sampling task (α = 0.92), and Destructive Effects of the Phobos Sampling task (α = 0.62). Scores from both scales were generally reliable except for the Destructive Effects of the Phobos Sampling tasks. Full items from the survey are detailed in the Appendix B. A complete timeline of the tasks and countermeasures survey is shown in Table 2.

Table 2. Timeline of each task and countermeasures survey within each quarter of the mission.

	Tasks	Surveys
Quarter 1: Day 1-11	**Rover** (Day 4) **Rover** (Day 6)	**Survey** (Day 5) **Survey** (Day 7)
Quarter 2: Day 12-22	**Rover** (Day 13) **Rover** (Day 19) **Rover** (Day 21)	**Survey** (Day 13) **Survey** (Day 20) **Survey** (Day 21)
Quarter 3: Day 23-34	**Rover** (Day 25) **Phobos** (Day 32) **Rover** (Day 33)	**Survey** (Day 25) **Survey** (Day 33) **Survey** (Day 33)
Quarter 4: Day 35-45	**Rover** (Day 35) **Phobos** (Day 38) **Rover** (Day 42)	**Survey** (Day 38) **Survey** (Day 38) **Survey** (Day 42)

By comparing responses from each crew during quarters where "recommended" pairings were implemented to responses during quarters when "worst" pairings were implemented, we assess whether implementing task pairings selected by the CREWS ABM can improve social relationships.

3. Results

3.1 Crew Member Experience During the Rover Task

Crew members responded to surveys about their experiences during the Rover task twice during the first quarter of the mission, three times during the second quarter, twice during the third quarter, and twice during the fourth quarter. Responses within each quarter were averaged in order to produce one observation per participant-quarter (n = 64).

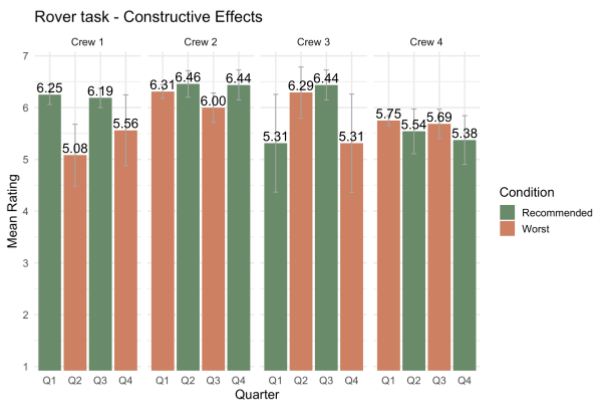

Fig. 6. Ratings of Constructive Effects of Rover pairings, by crew and by quarter. Means and standard error of survey responses in each period are depicted.

Figure 6 illustrates the average ratings of Constructive Effects for each of the four crews and each of the four quarters. During the first half of the mission (quarters 1 and 2) "recommended" pairings (M = 5.89, SD = 1.49) had a similar level of Constructive Effects for the Rover task compared to "worst" pairings (M = 5.86, SD = 0.96). The effect size of implementing "recommended" vs. "worst" pairings was assessed using Cohen's d [3]. Implementing "recommended" pairings had a negligible positive effect (d = 0.03) on reports of constructive relationships in the first half of the mission. In the second half of the mission, "recommended" pairings (M = 6.11, SD = 0.97) had a greater level of Constructive Effects on average compared to "worst" pairings (M = 5.64, SD = 1.50). Implementing "recommended" pairings had a small positive effect (d = 0.19) on reports of constructive relationships in the second half of the mission.

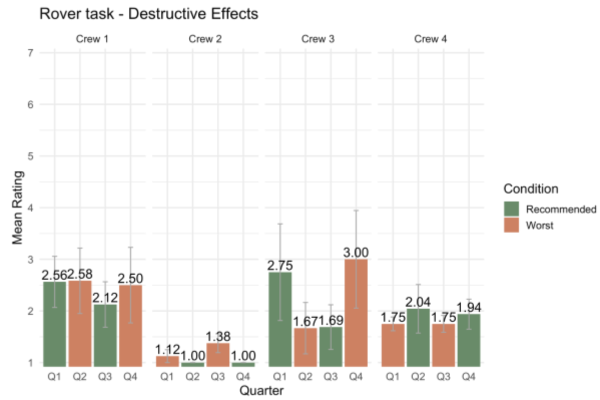

Fig. 7. Ratings of Destructive Effects of Rover pairings, by crew and by quarter. Means and standard error of survey responses in each period are depicted.

Figure 7 illustrates the average ratings of Destructive Effects for each crew and quarter. In the first half of the mission, "recommended" pairings (M = 2.09, SD = 1.50) had a higher level of Destructive Effects for the Rover task than "worst" pairings (M = 1.78, SD = 1.03). Implementing "recommended" pairings had a small positive effect (d = 0.25) on reports of destructive relationships in the first half of the mission. In the second half of the mission, "recommended" pairings (M = 1.69, SD = 0.80) had a lower level of Destructive Effects for the Rover task than "worst" pairings (M = 2.16, SD = 1.57). Implementing "recommended" pairings had a small negative effect (d = -0.39) on reports of destructive relationships in the second half of the mission.

We observe that the ABM "recommended" pairings succeeded in increasing constructive and decreasing destructive effects on the Rover task during the mission's second half. However, in the first half of the mission, "recommended" pairings had a negligible impact on constructive effects and increased "destructive" effects. A potential explanation for this is that the second half of the mission - in which crew members must contend with sleep deprivation, communication delay, and the extreme length of their social isolation - is a point of higher strain on crew social relationships. At this point, the task pairing countermeasures may have a more sizable effect on crew relationships. Additionally, the ABM may be incentivized to provide task pairing recommendations that work best during this period where social relationships are most strained.

3.2 Crew Member Experience During the Phobos Sampling Task

Crew members responded to surveys about their experiences during the Phobos Sampling task once during the third and fourth quarters of the mission. Responses in each quarter were averaged to produce one observation per participant-quarter (n = 32). Reliability was high for the scale on Constructive Effects of the Phobos Sampling task, and moderate for the scale on Destructive Effects of the Phobos Sampling task.

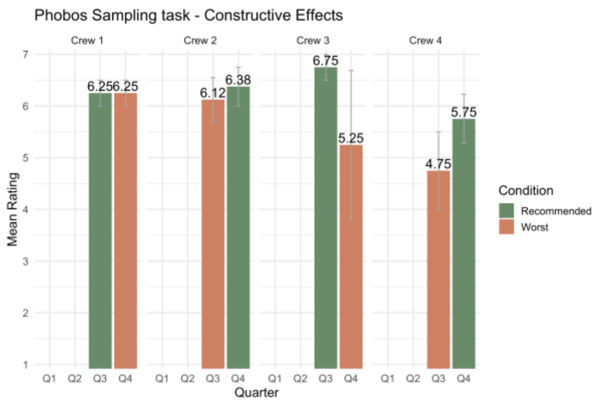

Fig. 8. Ratings of Constructive Effects of Phobos Sampling pairings, by crew and by quarter. Means and standard error of survey responses in each period are depicted.

Figure 8 illustrates the average ratings of Constructive Effects for each crew and quarter. The Phobos Sampling task was not performed during the first half of the mission. In the second half of the mission, "recommended" pairings (M = 6.28, SD = 0.78) produced higher ratings of Constructive Effects for the Phobos task than "worst" pairings (M = 5.59, SD = 1.65). Implementing "recommended" pairings had a medium positive effect (d = 0.56) on reports of constructive relationships. All four crews had equal or higher ratings of Constructive Effects under "recommended" pairings relative to "worst" pairings.

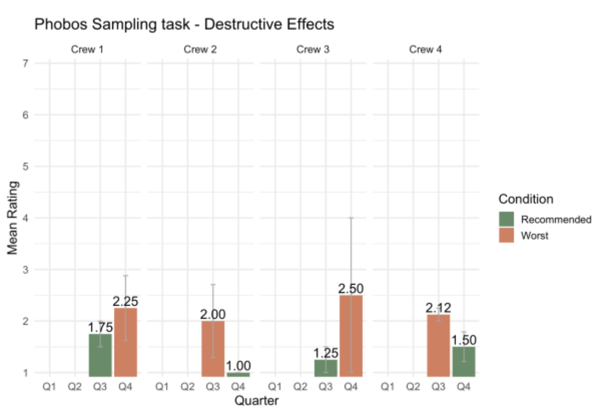

Fig. 9. Ratings of Destructive Effects of Phobos Sampling pairings, by crew and by quarter. Means and standard error of survey responses in each period are depicted.

Figure 9 illustrates the average ratings of Destructive Effects for each crew and quarter. When measured in the second half of the mission, "recommended" pairings (M = 1.38, SD = 0.50) produced lower ratings of Destructive Effects for the Phobos task than "worst" pairings (M = 2.22, SD = 1.60). Implementing "recommended" pairings had a medium negative effect (d = -0.73) on reports of destructive relationships. All four crews had lower ratings of Destructive Effects under "recommended" pairings relative to "worst" pairings. Overall, the "recommended" pairings succeeded in increasing constructive effects and decreasing destructive effects during the Phobos Sampling task.

3.3. Overall Social Relationships Between Crew Members

Next, we examine the effects of task pairings on generalized social ties, rather than items specific to the Rover and Phobos Sampling tasks. During each survey, participants responded to generalized questions about Task Affect ties ("With whom do you enjoy working?") and Hindrance ties ("Who makes tasks difficult to complete?"). In Tables 3 and 4, we report the proportion of time that crew members reported these ties as existing. Each participant responded about their ties towards the three other crew members - their partner during "recommended" task pairings, their partner during "worst" task pairings, and their middle partner, with whom they never worked with on tasks. These results summarize the frequency of which ties are held for each type of partner and are broken down by ratings of that partner during quarters where the "recommended" pairings were implemented and during quarters where the "worst" pairings were implemented.

Table 3. Comparison of frequency of Task Affect ties with partners under each condition

Proportion of Task Affect ties that exist between partners

	Task Affect ties felt towards ...			
	"Recommended" Partner	"Middle" Partner	"Worst" Partner	Total
During "Recommended" Pairing Quarters	1.00	0.94	0.94	0.96
During "Worst" Pairing Quarters	0.99	0.97	0.97	0.98
Total	0.99	0.96	0.96	

The results summarized in Table 3 suggest that crew members were more likely to report task affect ties with their recommended partner (0.99) than they were with their middle partner (0.96) or worst partner (0.96), regardless of whether "recommended" or "worst" pairings were implemented. Next, we examine whether working together in task pairings had a positive effect on relationships. We observe that crew members were only slightly more likely to enjoy working with their recommended partner when they were paired together on tasks (1.00) than when they were not paired together (0.99). However, crew members were more likely to enjoy working with their worst partner when they were paired together on tasks (0.97) as opposed to when they were not (0.94).

Table 4. Comparison of frequency of Hindrance ties with partners under each condition

Proportion of Hindrance ties that exist between partners

	Hindrance ties felt towards ...			
	"Recommended" Partner	"Middle" Partner	"Worst" Partner	Total
During "Recommended" Pairing Quarters	0.02	0.05	0.09	0.05
During "Worst" Pairing Quarters	0.02	0.05	0.04	0.03
Total	0.02	0.05	0.07	

From the hindrance ties summarized in Table 4, we can observe that regardless of whether "recommended" or "worst" pairings were implemented, crew members were less likely to report hindrance ties with their recommended partner (0.02) than they were with their middle partner (0.05) or their worst partner (0.07). Next, we examine whether working together on task pairings had a negative effect on relationships. Hindrance relationships with a crew member's recommended partners were not more or less likely during quarters where they were paired together on tasks (0.02) or quarters where they were not paired together (0.02). However, crew members were much less likely to hold hindrance relationships towards their worst partner when they were paired together on task (0.04) than when they were not paired (0.09).

Altogether, these pairings recommended by the model produced mixed results. As desired, crew members were more likely to form positive relationships with their recommended partners and more likely to form negative relationships with their worst partners. This suggests that the ABM was effective at predicting who would work well together, and that the ABM's recommendations could produce effective short-term pairings on a task.

We observed that placing individuals in their worst pairings has the potential to increase the frequency of positive relationships and decrease the frequency of negative relationships. This may be due to the type of performance tasks. Specifically, in most cases, crew pairings were able to successfully complete the assigned task. A shared goal on an interdependent task with a successful outcome could help improve the relationships. In terms of putting together pairs that have the best relationships, the "recommended" pairings were successful. However, implementing "worst" pairings was helpful in repairing problematic relationships between worst partners.

4. Discussion

Our findings provide partial evidence of the effectiveness of implementing task pairing countermeasures recommended with an ABM. "Recommended" pairings were higher rated by respondents in relation to the Rover and Phobos Sampling tasks. Participants assigned to "recommended" pairings reported more constructive effects and less destructive effects. Additionally, looking at overall social relationships outside of the tasks, participants were more likely to hold positive relationships (Task Affect) and less likely to hold negative relationships (Hindrance) with their "recommended" partners, relative to their worst partners. Overall, "Recommended" task pairings seem to have produced a more enjoyable experience for those working on the tasks.

A second question of interest is what can be done to prevent or repair negative relationships between crew members. When we compared differences between quarters where "recommended" and "worst" pairings were implemented, we found that pairing participants with their worst partners promoted task affect relationships and reduced hindrance relationships with these partners. Pairing participants with their worst partners they are less likely to get along with can improve their relationships with them. Likewise, pairing participants with recommended partners can improve their relationships with their recommended partners. However, this made smaller differences in the amount of task affect ties, likely because these relationships have less room to improve. Overall, this suggests that partnership on the Rover and Phobos Sampling tasks is beneficial for either recommended or worst partners. Forcing participants to work together, one-on-one, on highly interdependent tasks gives pairs time to focus on and improve their relationship. It repairs relationships between partners.

These findings pose an interesting tradeoff when selecting task pairings. "Recommended" pairings assembled participants who had better relationships and produced experiences on tasks that they viewed as more constructive and less destructive to their relationships. The "recommended" pairings placed effective partners together and were perceived more positively by participants. However, when it comes to non-task-specific network items, "worst" pairings helped participants improve their relationships with their task partners more than "recommended" pairings did. Even though tasks with worst partners were viewed as less constructive and more destructive to their relationships, in the long-term these interactions led to more task affect and fewer hindrance ties with worst partners. These destructive interactions between those who don't get along in the short-term may give partners an opportunity to repair their relationships in the long term.

These findings demonstrate a challenge in using empirically driven computational models to recommend countermeasures in social systems. The data used to develop and calibrate our model was gathered using participant surveys. Our ABM succeeded at recommending effective short-term pairings on the task that crew members thought would benefit their relationships. But in the long term, "worst" pairings provided a better chance to repair weak relationships within the crew, even if participants viewed the task as less constructive and more damaging in the short-term. This reveals a potential bias in building an ABM off survey responses, specifically towards recommending countermeasures that respondents view as most helpful, as opposed to helpful countermeasures that the respondents may not enjoy. These concerns should be considered when using survey-based ABMs to recommend countermeasures. Part of the concerns stem from ambiguity in the model's objective when selecting countermeasures - Is it better to exploit and strengthen the best relationships in a network? Or is it better to prioritize repairing the worst relationships? In the future, when using ABMs to recommend countermeasures, researchers should explicitly consider the tradeoff between supporting different types of ties in the network.

Our study also has several limitations. First, the limited number of participants that could be gathered in LDSE-analog studies constrained the sizes of both the sample used to develop/calibrate the ABM (eight crews), and the sample used to test potential countermeasures (four crews). An additional limitation of the study was the scope of the countermeasures. Within HERA, we re-assigned pairings for only two tasks out of a 45-day mission in which all crew members continuously worked and lived together. Because the decisions shaped by the ABM were small, only minor changes in social relationships among crew members were expected. However, the fact that we could observe small and medium effects on crew relations, despite the limited data to assist in ABM development and the limited scope of the pairing interventions being performed, is promising. These limitations should be kept in mind when interpreting the results, and the findings should only be viewed as preliminary support for the effectiveness of ABM-based countermeasures.

Our results provide a preliminary proof of concept for using prediction from an ABM to recommend countermeasures, in this case a task assignment and pairing strategy. There is a need to build on these findings with additional testing. Future work should develop models and test countermeasures in contexts where many participants are available. Contexts such as hack-a-thons or teams of remote workers, for instance, may provide an avenue for testing which imitates some of the characteristics of LDSE (e.g., isolation, sleep deprivation, slam shifts). Such contexts would provide large samples to establish more generalizable and reliable evidence of the effectiveness of ABM-recommended countermeasures, albeit with a lower fidelity approximation of what work in LDSE is like. The resulting findings could then be coupled with the small-sample high fidelity testing already performed in this paper.

Beyond space exploration, this paper demonstrates a generalizable approach for recommending decisions based on computational models. The use of computational models to preempt or mitigate deterioration of social relationships is relevant to teams and organizations back on Earth. In larger organizations that are unconstrained by the limited sample size of LDSE analogs, computational models could be built and tested at scale. Especially given the adoption of new technologies - *e.g.*, Zoom, Microsoft Teams, Slack, etc. - organizations now generate digital trace data that could be tapped to continuously calibrate and test such models in real time.

5. Conclusions

This study reports the results of developing an ABM to predict social relationships in LDSE analogs and the results of implementing countermeasures recommended by this ABM to improve social relationships. We observed that, by recommending pairs of two crew members to assign to certain tasks, our model was able to pair crew members with good relationships, producing more constructive and less destructive effects during the tasks.

Our findings show how the development of empirically validated ABMs has both *predictive* and *prescriptive* applications that support effective teamwork. Based on our research, we believe that continued development of ABMs has high potential for recommending countermeasures to support teams for space exploration, as well as to support teams back on Earth.

Acknowledgements

This research is based upon work supported by NASA under awards NNX15AM32G and 80NSSC18K0511. 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.

References

[1] B. Antone, A. Lungeanu, S. T. Bell, L. DeChurch, N. Contractor, Computational Modeling of Long-Distance Space Exploration: A Guide to Predictive and Prescriptive Approaches to the Dynamics of Team Composition, in: L. B. Landon, K. J. Slack, E. Salas (Eds.), Psychology and Human Performance in Space Programs, CRC Press, 2020, pp. 107–130.

[2] B. Antone, A. Gupta, S. Bell, L. DeChurch, N. Contractor, Testing Influence of Network Structure on Team Performance Using STERGM-Based Controls, Comput. 882 (2019) 1018 – 1030.

[3] S. S. Sawilowsky, New effect size rules of thumb, J Mod Appl Stat Methods. 8.2 (2009) 26.

[4] S. G. Hart, L. E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, Adv. Psychol. 52 (1988) 139–183.

[5] W. Arthur, B. D. Edwards, S. T. Bell, A. J. Villado, W. Bennett, Team task analysis: Identifying tasks and jobs that are team based, Hum. Factors. 47 (3) 654–669.

[6] W. Arthur, R. M. Glaze, A. Bhupatkar, A. J. Villado, W. Bennett, L. J. Rowe, Team task analysis: Differentiating between tasks using team relatedness and team workflow as metrics of team task interdependence, Hum. Factors. 54 (2) 277–295.

Appendix A (Social Network Survey)

Crew members were asked to nominate any of their fellow crew members in response to four sociometric items. Crew members were surveyed on these items throughout the course of the mission. Data collected on the first eight crew was used in the empirical calibration of the ABM. Crews 1 to 4 completed the social networks survey eight times over a 30-day mission. Crews 5 through 8 completed the social networks survey twelve times over a 45-day mission. Data collected from crews 9 through 12 was used to assess the effects of the crew re-pairing countermeasures implemented in these missions. Crews 9 through 12 completed the social networks survey on mission days 5, 7, 13, 20, 21, 25, 33, 38, and 42.

Task Affect ties: (M = 2.91 ties, SD = 0.28 ties)
"With whom do you enjoy working?"
Hindrance ties: (M = 0.15 ties, SD = 0.35 ties)
"Who made tasks difficult to complete?"
Leadership ties: (M = 2.69 ties, SD = 0.86 ties)
"To whom did you provide leadership?"
Followership ties: (M = 2.65 ties, SD = 0.61 ties)
"Who did you rely on for leadership?"

Appendix B (Countermeasures Survey)

The countermeasures survey was administered in-mission in order to assess the effects of "recommended" and "worst" pairings amongst the four crews in which they were implemented. The survey was completed by crews 9 through 12, on mission days 5, 7, 13, 20, 21, 25, 33, 38, and 42 of the mission. Because the Phobos Sampling task was only performed in the second half of the mission, the Phobos Sampling items were only included in the survey on mission days 20, 21, 25, 33, 38, and 42. Crew members were given the instructions "Please rate how you feel in response to the following statements." Responses were provided on a seven-point Likert-type scale, ranging from 1 = strongly disagree to 7 = strongly agree. Items were grouped into two scales, measuring constructive effects on relationships and destructive effects on relationships.

Rover Task Constructive Effects
($\alpha = 0.95$, $\mu = 5.87$, s.d. = 1.47)
- "Spending time with my partner on the Rover task was helpful to our relationship"
- "Working with my partner on the Rover task was a positive experience"

Rover Task Destructive Effects
($\alpha = 0.89$, $\mu = 1.92$, s.d. = 1.55)
- "Spending time with my partner on the Rover task was damaging to our relationship"
- "The Rover task added friction to my relationship with my partner"

Phobos Sampling Task Constructive Effects
($\alpha = 0.92$, $\mu = 5.94$, s.d. = 1.30)
- "Spending time with my partner on the Phobos task was helpful to our relationship"
- "Working with my partner on the Phobos task was a positive experience"

Phobos Sampling Task Destructive Effects
($\alpha = 0.62$, $\mu = 1.80$, s.d. = 1.24)
- "Spending time with my partner on the Phobos task was damaging to our relationship"

- "The Phobos task added friction to my relationship with my partner"

Appendix C (Task Attribute Survey)

All members of crews 5 through 12 completed a post-mission survey to assess the attributes of different categories of tasks completed throughout the HERA missions. All tasks in the HERA mission playbook were grouped into a total of 43 categories. This list of task categories was synthesized using the HERA playbook, task descriptions, and input of subject matter experts. Category names were selected to match those crew members would recognize from the schedules during HERA missions. For each of these categories, including the Rover task and Phobos Sampling task, crew members were asked items to rate attributes of that task category. This information was used to compare the Rover and Phobos Sampling tasks to other tasks completed within HERA.

NASA-TLX (Task Load Index) [4] measure of workload (α = 0.83):

Participants answered six items about each of the 43 different task categories:

- "How mentally demanding were each of the following tasks?"
- "How physically demanding were each of the following tasks?"
- "How hurried or rushed was the pace for each of the tasks?"
- "For each task, how hard did you have to work to accomplish your level of performance?"
- "How insecure, discouraged, irritated, stressed, and annoyed were you during each of the tasks?"
- "For each task, how successful were you in accomplishing what you were asked to do?"

Participants responded to each item on a 21-point scale from "very low" to "very high", identical to the original NASA-TLX. The six items were averaged, with the last item reverse-coded, to produce a single workload score for each task category.

TTA (Team Task Analysis) [5,6] measure of task interdependence:

Participants were given the prompt: "For each of the tasks you have worked on over the course of the HERA mission, please rate the extent to which you were required to work with other crew members for optimal performance." They were asked to answer the item about each of the 43 different task categories. The single-item measure was assessed on a five-point Likert-type scale, ranging from 1 = "Not at all", 2 = "Very little", 3 = "Somewhat", 4 = "Quite a bit", and 5 = "Very much".

The complete list of 43 task categories that crew members responded about is provided below:

Asteroid Sample Analysis; Blood Draw; Body Measurement Equipment Use; Brine Shrimp; Cognition/Luminosity Survey Brain Games Activity; Crew Down Time; Daily Diary; Daily Planning Conference; Daily Prep; Egress; Emergency Simulation; Exercise; Fatigue Interface Testing; Hardware Don/Doff; HERA Module Equipment Maintenance; Housekeeping; Hydrolysis Module Build/Swap Out; Hygiene; Light Measurement using the light meter; LiOH Canister Change Out; Low Latency Teleoperations Team Task; Meal w/ ISS FIT; Medical Exam; Microbiome Testing; MiniPCR; MMSEV-EVA; Multiteam Task Battery; On Board Training

PAO (Public Affairs Office) Event; Potable Water Check; Private Family Conference; Private Medical Conference; Private Psychological Conference; Psychomotor Vigilance Task (visual stimulant reaction task); Rover; S-COG (Stahn's Cognition and Cognitive Test Battery); Seeds; Self-Scheduling; Simulator Platform (flight simulators); Surveys; Systems Maintenance Simulation; Systems Status Check (Team Simulation Task); Team Task Battery.