Fiscal Year:	FY 2024	**Task Last Updated:**	FY 01/31/2024
PI Name:	Chaspari, Theodora Ph.D.		
Project Title:	Artificial Intelligence for Tracking Micro-Behaviors in Longitudinal Data and Predicting Their Effect on Well-Being and Team Performance		
Division Name:	Human Research		
Program/Discipline:			
Program/Discipline--Element/Subdiscipline:			
Joint Agency Name:		**TechPort:**	Yes
Human Research Program Elements:	(1) **HFBP**:Human Factors & Behavioral Performance (IRP Rev H)		
Human Research Program Risks:	(1) **Team**:Risk of Performance and Behavioral Health Decrements Due to Inadequate Cooperation, Coordination, Communication, and Psychosocial Adaptation within a Team		
Space Biology Element:	None		
Space Biology Cross-Element Discipline:	None		
Space Biology Special Category:	None		
PI Email:	chaspari@tamu.edu	**Fax:**	FY
PI Organization Type:	UNIVERSITY	**Phone:**	979-458-2205
Organization Name:	Texas A&M Engineering Experiment Station		
PI Address 1:	3112 Tamu		
PI Address 2:			
PI Web Page:			
City:	College Station	**State:**	TX
Zip Code:	77843-0001	**Congressional District:**	17
Comments:			
Project Type:		**Solicitation / Funding Source:**	2020 HERO 80JSC020N0001-FLAGSHIP, OMNIBUS1 Human Research Program: Crew Health Appendix A; Omnibus1-Appendix B
Start Date:	03/09/2022	**End Date:**	12/31/2023
No. of Post Docs:	0	**No. of PhD Degrees:**	1
No. of PhD Candidates:	0	**No. of Master' Degrees:**	0
No. of Master's Candidates:	0	**No. of Bachelor's Degrees:**	0
No. of Bachelor's Candidates:	0	**Monitoring Center:**	NASA JSC
Contact Monitor:	Whitmire, Alexandra	**Contact Phone:**	
Contact Email:	alexandra.m.whitmire@nasa.gov		
Flight Program:			
Flight Assignment:	NOTE: End date changed to 12/31/2023 per NSSC information (Ed., 7/6//23) NOTE: End date changed to 08/31/2023 per A. Beitman/JSC (Ed., 2/22/23) NOTE: End date changed to 03/08/2024 per NSSC information (Ed., 2/15/23)		
Key Personnel Changes/Previous PI:	NOTE: Per the Principal Investigator (PI): Dr. Roma and Dr. Loerch are no longer with the project (Ed., 7/6/23).		
COI Name (Institution):	Bell, Suzanne Ph.D. (NASA Johnson Space Center)		
Grant/Contract No.:	80NSSC22K0775		
Performance Goal No.:			
Performance Goal Text:			

Task Description:	Future long-distance space exploration will have a number of challenges that increase the risk of inadequate cooperation, coordination and psychosocial adaptation, and can lead to behavioral health and performance decrements. In NASA-sponsored analogs, the primary methodology for capturing team interaction data is self-report surveys. While this method may provide some insights, it has significant limitations and biases. We hypothesize that micro-behaviors detected by artificial intelligence (AI) can provide unique insights into emotional reactivity and operationally-relevant team performance beyond self-report team functioning measures commonly used in NASA-funded research. Our research has three primary aims: (1) Leverage advanced multimodal data analytics to detect micro-behaviors, including negative micro-behaviors (e.g., acidic humor, sarcastic/caustic comment, underestimation of the other's ideas or ability, rudeness and insensitivity, fleeting minute-by-minute disagreements) and positive micro-behaviors (e.g., actions that reflect inclusion and caring, active listening, recognizing others' achievements, using friendly expressions and tone of voice) in longitudinal team interactions; (2) Identify physiological reactivity to micro-behaviors; and (3) Incorporate knowledge on micro-behaviors to predict team performance. To address these aims, we leveraged previously collected self-reported surveys, physiological, and audio data from Human Exploration Research Analog (HERA) Campaigns 4 and 5 via NASA grant NNX16AQ48G (PI = Bell) that focused on team composition and interpersonal dynamics, and the NASA Human Factors and Behavioral Performance Exploration Measures project (HFBP-EM; PI = Bell). The data were collected from 9, 4-person teams (n = 36, 11 women, 11 with military background) in isolation and confinement for up to 45 days. Each crew conducted five team interaction battery (TIB) tasks, one during the pre-mission phase and four during the in-mission phase, in which the crew performed two types of tasks (i.e., decision-making task, relational task). In this study, we used the audio and transcribed data collected during the TIB. There were a total of 44 TIB interactions and each interaction lasted an average of 1.5 hours. Additionally, daily team performance data were collected under HFBP-EM (PI = Bell). Crew members rated their perception of their crew's performance using a 7-point Likert scale in which higher numbers reflected higher performance across four items (e.g., "To what extent did the crew accomplish its primary goals today?" "To what extent were the important tasks for today done in a high-quality fashion?"). On days of the TIB, team performance was operationalized as the mean of the crew member's ratings of team performance on that day. Our research products contribute to reducing the Team Risk, particularly gaps 102 and 106. Identified key micro-behaviors that affect well-being and team functioning can be used as unobtrusive measures with which to monitor team functioning. Insights from this project can inform targeted personalized pre-mission and in-mission intervention strategies (e.g., micro-video training) that suggest concrete action items to crew members and gradually adapt recommendations for a specific person and/or team.
Rationale for HRP Directed Research:	
Research Impact/Earth Benefits:	Results from this research contribute to a greater understanding of factors that affect team operation and performance in isolated and confined environments (ICE), and the effective management of future space crews. Particularly relevant to this project is the focus on micro-behaviors, that is small, often unconscious gestures, words and tone of voice which can influence how included (or not included) the people in the team feel. The proposed AI methods allow us to track micro-behaviors on a moment-to-moment basis, thus potentially yielding a more complete picture of team dynamics. This will help to tease out the micro-behaviors that matter most (even in an unconscious manner), and to identify protective factors against negative micro-behaviors. The type and severity of each micro-behavior and the corresponding emotional reactivity measures will further contribute in combination with existing self-report measures of relational and team functioning to predict operational team performance. The specific research questions and methodologies are developed and tested on HERA data and are specific to ICE; thus, beyond space crews, findings and systems developed as part of this research can be applied to Earth teams that operate in ICE such as expedition and science teams in the Arctic and Antarctic.
	Task 1: Leverage advanced multimodal data analytics to detect positive and negative micro-behaviors in longitudinal team interactions This task was divided into three parts: (1) We designed a coding manual and coded the discussions of the Team Interaction Battery (TIB) task on micro-behaviors; (2) We built conversational markers of micro-behaviors that take into account linguistic and acoustic measures of dialogue (in)coherence, (im)polite language, empathy and sarcasm; and (3) We designed context-dependent machine learning models that take as an input the aforementioned linguistic and acoustic markers, and consider the context of the dialog between members (i.e., turn-taking, underlying dialog sentiment) for automatically detecting micro-behaviors from speech. Task 1.1: Coding micro-behaviors in team interactions To capture negative and positive micro-behaviors in the TIB task, we utilized coded data from NASA Grant NNX16AQ48G (PI: Bell). This coding scheme incorporated Gushin and colleagues' (2016) analysis of space crew communications with the Discussion Coding System (Schermuly & Scholl, 2012) to capture the functional meaning of different communication exchanges (e.g., initiate joint activity, regulate conversation). Transcripts of the TIB were coded at the turn level across five main functions followed by a secondary function. Agreement was high (Kappa >.80) for all categories. Given that the prior data were coded at the turn level, we were able to investigate micro (i.e., moment-by-moment) interactions as part of the current research project. To meet this objective, our team re-classified the behavioral functions as positive (i.e., enhances team functioning), negative (i.e., degrades team functioning), or neutral micro-behaviors. Task 1.2: Extracting linguistic and acoustic markers of micro-behaviors We extracted acoustic and linguistic features for each dialog turn. These features have been found indicative of emotional arousal, (im)politeness, empathy, and sarcasm and have been widely used in prior work that aims to model the nature of human interactions (Cheang & Pell, 2008; Otterbacher et al., 2017). Lexicon-based features were extracted based on two dictionaries, namely, the Linguistic Inquiry and Word Count (LIWC) (Tausczik et al., 2010) and STRESSnet (Driskell et al., 2023). Apart from these interpretable linguistic features, we also utilized state-of-the-art feature extraction techniques in natural language processing (NLP) to compute word embeddings from the transcripts. We used the paragram embedding (Wieting et al., 2015), which is trained on the Paraphrase Database, that takes into account paraphrase pairs to accurately model semantics. We extracted frame-based acoustic features of the root-mean-square energy, zero-crossing rate, voicing probability, fundamental frequency (F0), and 12 Mel-frequency cepstral coefficients (MFCC). To obtain the turn-level features, we computed a set of statistics (i.e., minimum, maximum, range, absolute value of the maximum and minimum, mean, slope, offset of linear approximation, quadratic error, standard deviation, skewness, kurtosis) of the frame-based features along with their smoothed version over each turn. We extracted 4 lexicon-based features corresponding to positive, negative, neutral, and compound sentiments of the target participants via the Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto & Gilbert, 2014) algorithm, to measure the underlying sentiment

in the conversation. Turn-similarity features were calculated by measuring the cosine similarity score between the sender and the target speaker with respect to their acoustic and linguistic features in order to capture convergent or divergent language or tone of voice between interlocutors.

Exploratory analysis suggests that negative micro-behaviors included increased sentences with passive voice for both the sender and the target, potentially due to the subtlety of their expression. Negative micro-behaviors also depicted fewer second-person personal pronouns (e.g., “you”) compared to turns without those behaviors. While this finding for negative micro-behaviors might initially be nonintuitive, since blaming language tends to include a large number of second-person pronouns, this result might also reflect the fact that negative micro-behaviors are many times indistinct and not sharply expressed. Positive micro-behaviors had significantly higher positive effects, a lower percentage of angry language, and a higher percentage of language related to achievements compared to negative micro-behaviors. Acoustic turn similarity between sender and target was the highest for instances of positive micro-behaviors (i.e., indicating convergence between speakers) and the lowest for instances of negative micro-behaviors (i.e., indicating divergence between speakers).

Task 1.3: Context-dependent machine learning models for the automatic detection of micro-behaviors We leveraged machine learning classifiers and dialog state tracking models, combined with natural language processing techniques relying on lexicon-based methods and data-driven methods, to learn complex multimodal patterns of micro-behaviors and automatically detect positive and negative micro-behaviors between team members. We used various combinations of the interpretable feature groups (i.e., lexicon-based, acoustic, turn similarity) to train a random forest (RF) classifier that classified between the three considered classes (i.e., micro-aggression, micro-affirmation, none). We also examined four variations of deep learning models that rely on long short-term memory (LSTM) neural networks and dialog state tracking systems. Our experimental findings indicate that the psycholinguistic markers extracted using the linguistic inquiry and word count (LIWC), STRESSnet dictionaries, and acoustic features can achieve an f1-score up to 55% in a three-class classification problem, which is well above the 33% chance accuracy. Our findings also suggest that modeling turns between the sender and target of micro-behaviors is significantly more effective in detecting micro-behavior than only modeling the sender’s information. Finally, we demonstrate the effect of introducing context for detection purposes. Dialog state tracking approaches that model the linguistic interaction between team members and incorporate contextual information about the task and sentiment of the conversation can further yield improved performance, depicting an f1-score of 57.73%. These results suggest the preliminary feasibility of such a system using unobtrusive speech measures. Training these large-scale ML models with more data (e.g., additional mission tasks) can potentially yield improved results and increase the robustness of these automated systems.

Task 2: Identify physiological reactivity to micro-behaviors

Physiological reactivity to micro-behaviors was measured in order to better understand how a target of micro-behavior experiences the focal event. Physiological reactivity was measured in terms of heart rate (HR), for which continuous data from the TIB task of the C4 campaign was captured. Physiological reactivity was defined as the average heart rate in the time interval before the micro-behavior occurrence subtracted from the average heart rate in the interval during/after the micro-behavior occurrence, i.e., Reactivity = Average HR (During/After) – Average HR (Before). Physiological reactivity was computed for positive and negative micro-behaviors separately for each team member.

Task Progress:

Results indicate a large inter-individual variability in terms of the physiological reactivity responses. A subset of team members depicted increased physiological reactivity to negative micro-behaviors and decreased physiological reactivity to positive micro-behaviors, which is anticipated. However, other team members, such as P4, P5, and P6, depicted opposite trends. These results suggest that different team members might depict different sensitivity to micro-behaviors. Future analysis should also acknowledge and take into account additional confounding factors that might affect physiological reactivity (e.g., physical activity, daily stressors) and other buffering mechanisms.

Task 3: Incorporate knowledge on micro-behaviors to predict team functioning

In this task, (1) we investigated the dependence of team performance on subtle linguistic expressions between team members and (2) examined the extent to which ML models of team interactions that rely on language features can effectively predict team performance degradation.

Task 3.1: Investigating the dependence of team performance on subtle linguistic expressions between team members We examined to what extent the linguistic content of a micro-behavior was associated with team performance outcomes. We studied the dependence of team performance on linguistic features extracted using the Linguistic Inquiry and Word Count (LIWC) dictionary, which captures affective, cognitive, and physical processes, as well as drives and motives, extracted from conversation turns corresponding to positive and negative micro-behaviors through a linear mixed-effect (LME) model. LIWC features included: (1) two measures of affective processes, including positive and negative emotion; (2) three measures of cognitive processes, that included discrepancy (e.g., would, can, want, could), certainty (e.g., really, actually, of course), differentiation (e.g., but, not, if, or); (3) three measures that capture physical processes related to food, body, and sex; and (4) five measures that capture drives and motives, including affiliation (e.g., we, our, us, help), achievement (e.g., work, better, best, working), power (e.g., own, order, allow, power), reward (e.g., opportunity, win, gain, benefit), and risk (secure, protect, pain, risk). These measures were chosen in an empirical and theoretical manner since we anticipated that micro-behaviors, as they were coded, would include content relevant to these processes and dimensions.

Results indicated that increased use of language associated with positive emotion during positive micro-behaviors is associated with an increase in team performance. On the contrary, increased use of negative emotion language during negative micro-behaviors is associated with a decrease in performance. When running the same LME model using data from all turns, the language associated with positive emotion appears to be positively associated with team performance reaching almost significance level, but the language with negative emotion does not depict a significant association with team performance. This potentially suggests that the affective content of micro-behaviors can be an indicator of team performance, but the affective content present over the entire conversation does not necessarily predict team performance, thus highlighting the importance of studying team interaction in the context of micro-behaviors. Language associated with discrepancy is negatively associated with team performance, potentially because discrepancy might indicate disagreement that cannot be resolved. Language associated with certainty is positively associated with team performance since certainty might indicate a high degree of confidence of the team about environmental assessments and subsequent decisions. Language related to food expressed during positive micro-behaviors depicts a significantly positive correlation with team performance, while language related to sex during negative micro-behaviors is significantly negatively correlated with team performance. Finally, reward-related language when computed over all turns depicts a significant positive association with improved team performance, a finding which is anticipated since this type of language refers to gains and benefits and is likely to co-occur with high performance.

Task 3.2: Predicting team performance degradation using language features Second, we propose an approach to model the dynamics of team interaction and the temporal nature of this interaction for detecting degradation in team performance. To understand changes in team performance, we binarized the outcome such that if the crew's average score for a given day fell below the within-crew mean performance across the mission, the team was classified as having a low-performing day. If it was higher than the team's average, the day was classified as high-performing. In our proposed approach, we simultaneously capture the team interaction via graph neural networks (GNN) and the temporal sequence of the conversation via recurrent neural networks (RNN) to learn linguistic embeddings that are indicative of high or low team performance. The proposed model can capture essential components of team interaction, such as the direction of the conversation between team members and the history of conversations across days (i.e., average directed interaction among a pair of team members). We used the Sentence-BERT embedding as the turn-based vector representation. Results suggested that modeling the interaction among team members and the flow of the conversation together helps in achieving significantly higher performance ($p<0.01$), with 64.05% unweighted recall compared to the 50% chance level for binary classification, compared to modeling each of the factors separately.

Summary of Findings and Implications to Space Missions The preliminary analysis conducted in this project indicates that by leveraging speech-based data analytics, NASA could gain deeper insights into the dynamics of astronaut teams. Important findings that can be considered for future space missions are the following: (1) Negative micro-behaviors are characterized by increased use of passive voice, an increase in vocal pitch (F0), and decreased vocal similarity between interlocutors (Task 1 analysis); (2) Positive micro-behaviors are characterized by decreased use of passive voice, an increase in language related to positive affect, and achievement, and increased vocal similarity between interlocutors (Task 1 analysis); (3) Language related to positive affect, certainty, food, and reward predict an increase in team performance (Task 3 analysis); (4) Language related to negative affect, discrepancy, and sex predict a decrease in team performance (Task 3 analysis); and (4) Inter-individual variability was observed in the way micro-behaviors are physiologically experienced (Task 2 analysis). Rather than a theory-driven approach, which is difficult to operationalize based on the literature from domains less relevant to space missions, our empirical results can inform pre-mission and in-mission interventions. For instance, fostering positive communication among crew members could be pivotal in enhancing overall mission success. Specific linguistic cues related to positive affect, cognition, unbiased physical referents, and positive motives are areas where interventions can be targeted. NASA could further apply predictive AI models developed here to alternate sources of data, such as the ones from Artemis and depending on anticipated crew adoption, to unobtrusively track team behaviors and predict decrements in performance in a personalized manner without relying on additional sensors, thus enhancing crew privacy.

Bibliography Type:	Description: (Last Updated: 03/07/2024)
Abstracts for Journals and Proceedings	Begerowski SR, Khader AM, Paromita P, Chaspari T, Bell ST. "Through thick and thin: Predicting team performance using history of interaction patterns via deep learning models." 39th Annual Society for Industrial Organizational Psychology (SIOP) Conference, Chicago, Illinois, April 17-20, 2024. Abstracts. 39th Annual Society for Industrial Organizational Psychology (SIOP) Conference, Chicago, Illinois, April 17-20, 2024. , Apr-2024
Abstracts for Journals and Proceedings	Paromita P, Khader A, Begerowski SR, Bell ST, Chaspari T. "Speech-based artificial intelligence for detecting micro-behaviors and predicting their effect on team performance." 2024 NASA Human Research Program Investigators' Workshop, Galveston, Texas, February 12-14, 2024. Oral presentation. 2024 NASA Human Research Program Investigators' Workshop, Galveston, Texas, February 12-14, 2024. , Feb-2024
Abstracts for Journals and Proceedings	Paromita P, Khader A, Begerowski SR, Bell ST, Chaspari T. "A linguistic analysis on the impact of team interactions on team performance during space exploration missions." 2024 NASA Human Research Program Investigators' Workshop, Galveston, Texas, February 12-14, 2024. Poster presentation. 2024 NASA Human Research Program Investigators' Workshop, Galveston, Texas, February 12-14, 2024. , Feb-2024
Abstracts for Journals and Proceedings	Paromita P, Khader A, Begerowski S, Bell ST, Chaspari. "A linguistic analysis on the impact of team interactions on team performance during space exploration missions." 2024 IEEE International Conference on Affective Computing and Intelligent Interaction (ACII). To be submitted. 2024 IEEE International Conference on Affective Computing and Intelligent Interaction (ACII). , Mar-2024
Abstracts for Journals and Proceedings	Begerowski SR, Khader AM, Paromita P, Chaspari T, Bell ST. "What's that supposed to mean? Capturing micro-behaviors in teams." 38th Annual Society for Industrial Organizational Psychology (SIOP), Boston, Massachusetts, April 19-22, 2023. Abstracts. 38th Annual Society for Industrial Organizational Psychology (SIOP), Boston, Massachusetts, April 19-22, 2023. , Apr-2023
Abstracts for Journals and Proceedings	Paromita P, Khader A, Begerowski SR, Bell ST, Chaspari, T. "Linguistic and vocal markers of micro-behaviors between team members during analog space exploration missions." 2023 NASA Human Research Program Investigators' Workshop, Galveston, Texas, February 7-9, 2023. Poster presentation. 2023 NASA Human Research Program Investigators' Workshop, Galveston, Texas, February 7-9, 2023. , Feb-2023
Books/Book Chapters	Paromita P, Khader A, Begerowski S, Bell ST, Chaspari T. "Speech descriptors of micro-behaviors during team interactions." in "Computational Group and Team Dynamics: Forging an Interdisciplinary Science. " Ed. Steve W. J. Kozlowski, Hayley Hung, Nale Lehmann-Willenbrock, Albert Ali Salah. Oxford, England: Oxford University Press, 2024. (In progress as of March 2024)., Aug-2024
Dissertations and Theses	Paromita P. "Fostering a healthier work environment: An artificial intelligence-based approach for identifying individual and team well-being in real-world settings. " PhD Thesis, Texas A&M University, December 2023. , Dec-2023