Fiscal Year:	FY 2021	**Task Last Updated:**	FY 02/28/2021
PI Name:	Lee, John Ph.D.		
Project Title:	HCAAM VNSCOR: Conversation Analysis to Measure and Manage Trust in Virtual Assistants		
Division Name:	Human Research		
Program/Discipline:			
Program/Discipline--Element/Subdiscipline:			
Joint Agency Name:		**TechPort:**	No
Human Research Program Elements:	(1) **HFBP**:Human Factors & Behavioral Performance (IRP Rev H)		
Human Research Program Risks:	(1) **HSIA**:Risk of Adverse Outcomes Due to Inadequate Human Systems Integration Architecture		
Space Biology Element:	None		
Space Biology Cross-Element Discipline:	None		
Space Biology Special Category:	None		
PI Email:	jdlee@engr.wisc.edu	**Fax:**	FY
PI Organization Type:	UNIVERSITY	**Phone:**	608-890-3168
Organization Name:	University of Wisconsin, Madison		
PI Address 1:	Department of Industrial and Systems Engineering		
PI Address 2:	1513 University Ave		
PI Web Page:			
City:	Madison	**State:**	WI
Zip Code:	53706-1539	**Congressional District:**	2
Comments:			
Project Type:	Ground	**Solicitation / Funding Source:**	2017-2018 HERO 80JSC017N0001-BPBA Topics in Biological, Physiological, and Behavioral Adaptations to Spaceflight. Appendix C
Start Date:	04/15/2019	**End Date:**	04/14/2023
No. of Post Docs:		**No. of PhD Degrees:**	
No. of PhD Candidates:	1	**No. of Master' Degrees:**	
No. of Master's Candidates:		**No. of Bachelor's Degrees:**	
No. of Bachelor's Candidates:	1	**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 per S. Huppman/HRP and NSSC information (Ed., 3/20/2020) NOTE: End date changed to 3/31/2020 per NSSC information (Ed., 1/22/2020)		
Key Personnel Changes/Previous PI:	March 2021 report: Dr. Kerry McGuire is no longer working on the project; Jerri Stephenson is now CoInvestigator.		
COI Name (Institution):	Cross, Ernest Ph.D. (NASA Johnson Space Center) Stephenson, Jerri M.S. (NASA Johnson Space Center)		
Grant/Contract No.:	80NSSC19K0654		
Performance Goal No.:			
Performance Goal Text:			

Task Description:

This task is part of the Human Capabilities Assessments for Autonomous Missions (HCAAM) Virtual NASA Specialized Center of Research (VNSCOR).
The goal of this research is to develop conversation analysis to measure and mitigate inappropriate trust in virtual assistants. These trust measurements will guide system design, particularly the multimodal interactions and mode switching, as well as how to mitigate over trust and trust recovery. We will use conversation analysis to measure trust at multiple time-scales from real-time interactions to longitudinal monitoring of trust over a long duration exploration mission.

Conversation analysis provides a promising, but relatively unexplored approach to measuring trust. We propose a conversation analysis at the micro, meso, and macro levels which includes not just the words, but also pauses and facial expressions. Specifically, at the micro-level, conversation elements include voice inflections, pauses between works and keystrokes, gaze shifts, and facial expressions. The meso-level analysis includes words exchanged during interactions with the virtual assistant along with other team interactions as they relate to the automation. At the macro level, conversational analysis considers interaction time, interaction effort, frequency of interaction, turn-taking, barging in tendency, and whether it is the person or the virtual assistant who initiates the interaction. Additionally, prior research into conversational analysis indicates there are novel ways of managing or calibrating trust through the presentation of information, e.g., manipulating the tone and cadence of the system when using speech and through facial expressions (Nass & Brave, 2005; DeSteno et al., 2012).

Due to time delays in communication, long duration exploration missions will require greater crew autonomy and greater reliance on automation. For this approach to work trust calibration needs to be engineered into the system. Trust is a critical construct that mediates how well human operators use automated systems, such as virtual assistants, that provide decision support. Trust affects people's willingness to rely on automated systems in situations that have a degree of uncertainty and risk. Trust strongly affects the effectiveness of human-agent collaboration, particularly in the willingness to accept suggestions from a virtual assistant. Knowing whether or not to trust automation can be further complicated by lack of sleep, workload, task risk, and task complexity. Moreover, as we continue to push the limits of intelligent systems and rely on them more as decision aids trust calibration (i.e., operator trust is at a level which matches the automation's capabilities) becomes essential to mission execution.

Appropriate calibration of trust requires matching the operator's trust to the virtual assistant's current capabilities. Calibration of trust is not something that can happen once, but must occur throughout the life cycle of the interaction between operator and automated system (Hoffman et al., 2009). Trust is a dynamic construct that continuously increases and decreases due to a number of factors, primary the performance of the automated system, i.e., higher performance leads to higher trust and vice versa. Although much effort focuses on creating more capable and trustworthy automation, less effort has considered the equally important consideration of creating trustable automation. Trustable automation is automation that is understandable and that naturally promotes calibrated trust. Therefore, we aim to create trustable automation by continuously measuring operators' trust unobtrusively and in real-time, and then use this measure to guide the virtual agent to employ one or more countermeasures to calibrate trust and improve human-system performance.

References

DeSteno D, Breazeal C, Frank RH, Pizarro D, Baumann J, Dickens L, Lee JJ. Detecting the trustworthiness of novel partners in economic exchange. Psychol Sci. 2012 Dec;23(12):1549-56. http://doi.org/ ; PubMed PMID: 23129062

Hoffman RR, Lee JD, Woods DD, Shadbolt N, Miller J, Bradshaw JM. The dynamics of trust in cyberdomains. IEEE Intelligent Systems. 2009 Nov-Dec;24(6):5-11. https://

Nass C, Brave S. Wired for Speech : How Voice Activates and Advances the Human-Computer Relationship. Cambridge, MA: MIT Press, 2005.

Rationale for HRP Directed Research:

Research Impact/Earth Benefits:

The outcomes of this research will make two important contributions to the overall HCAAM VNSCOR effort. First, it will promote more effective interactions and acceptance of virtual assistants. Second, it will provide new analytic techniques for understanding how people work with automated agents as team members.
Virtual assistants and other types of agents enabled by artificial intelligence represent an important opportunity to extend human capabilities, but only if they are accepted and trusted appropriately. If people trust the virtual assistant too much they will rely on it in situations that exceed its capability, and if they trust it too little they will fail to engage it when it could benefit the team. One pathway towards appropriate trust is to make the virtual assistant more trustworthy: increase its technical capabilities to accommodate any situation. Another approach is to make it more trustable: communicate its capability and allow its capability to be challenged in its interactions with people. Such trustable technology requires three important advances to the state of knowledge in the field:

1. An ability to ascertain how much people currently trust the technology

2. An ability to convey uncertainty and its capability, particularly as part of conversational interactions

3. Interaction affordances that provide the opening for people to assess the capability of the assistant, particularly as part of conversational interactions.

These three advances for trustable technology require the development of new analytic techniques for understanding human interaction with automated teammates. Real-time, unobtrusive measures of trust represent a particularly valuable, but challenging measure to develop. Trust is most often measured with ratings and indirectly through people's decision to rely on automation, which are obtrusive not diagnostic. Conversation and text-based interactions offer a promising, but unexplored way to assess trust. Text analysis has a 50-year history in domains as diverse as psycholinguistics and cognitive science, and more recently natural language processing, affective state assessment, and sentiment analysis. Building on the foundation of text analysis makes it possible for this research to immediately contribute to data analysis of previous and future studies of automation-human teaming, and to contribute to the foundation of conversational agent design.

Task Progress:	The major goals of the project during this phase: * Developed the PRocedure Integrated Development Environment (PRIDE) procedure microworld and conversational agent * Review and integrate subjective rating scales of trust to guide the selection of scales and to create a trust lexicon * Develop a conceptual framework to guide on interaction with intelligent agents The major activities associated with these goals include: Habitat maintenance system testbed: We developed adapted electronic procedure software to the task of maintaining the International Space Station habitat systems. The agent preprogrammed with the system layout and procedure protocols can provide assistance and recommendations for participants to follow and operate the procedures of maintaining the habitat in the PRIDE system. Specifically, participants need to remove carbon dioxide using the Carbon Dioxide Removal System (CDRS), actively cool devices using the Active Thermal Control System (ATCS), and distribute the power supply using the Electrical Power System (EPS). These three systems are interdependent: EPS distributes power generated from the solar arrays to both CDRS and ATCS. ATCS provides cooling for CDRS. The CDRS blows air from the cabin across beds that are heated to remove humidity and absorb carbon dioxide. The absorbed carbon dioxide is then released and vented to space. The scrubbed air is cooled and humidified by the water supplied from ATCS before returning to the cabin. Participants will be asked to control the habitat system for removing the CO2 from the air, which requires to control these three systems in a specific order: verify power fuse boxes in EPS to ATCS, configure heat exchanger for cooling, start up the ATCs, verify power fuse boxes in EPS to CDRS, switch CDRS modes, and record values after the activation. While the habitat maintenance task is being operated automatically using the PRIDE system, the participants should engage in the secondary task on system status checking using conversational agent as a platform for the secondary task. Since such communication would be a common practice for future missions, this secondary task holds a high external validity, which means it can be well-generalized to an operational setting. A conversational agent provides an unobtrusive and natural way to measure trust. Compile and integrate trust scales. Trust has emerged as a prevalent construct to describe relationships between people and between people and technology in myriad domains. Across disciplines and application domains, researchers have relied on many different questionnaires to measure trust. The degree to which these scales differ has not been systematically explored. We used a word-embedding text analysis technique to identify the differences and common themes across the most commonly used trust questionnaires and provide recommendations for questionnaire selection. A mapping review was first conducted to identify the existing trust questionnaires. In total, we included 40 trust questionnaires from three main domains (i.e., Automation, Humans, and E-commerce) with a total of 506 items measuring different dimensions/types of trust (i.e., Dispositional, History-based, and Situational). Next, we encoded the words within each questionnaire using GloVe word embeddings and computed the embedding for each questionnaire item, and for each questionnaire as a whole. We reduced the dimensionality of the resulting dataset using UMAP (Uniform Manifold Approximation and Projection) to visualize these embeddings in scatterplots. The scatterplots show which questionnaires, items, and words are close to each other. This semantic space shows how trust has been operationalized, serves to produce a lexicon of trust-related words, and also guides questionnaire selection based on domain and trust dimensions. Overall, word embedding provides a novel means to compare trust scales. Review literature of trust between interacting agents. We completed a major literature review to develop a conceptual framework of trust between agents. Some of these components of this review include the socio-technical system factors of the goal environment, such as organizational structures (e.g., chain of command, management practices, incentives); time constraints of the task; assigned or resulting workload; individual preferences; and perceived risk. The type and length of sequences directly influence the structure of subsequent situations, which influences the strategy that an agent may use. Strategy can draw from outside knowledge, but is embedded within the goal environment, and can be constrained by the situation structure and sequence. Semiotics may start with an interface and its associated design artifacts (e.g., observable signals, display design features), but it also includes the agents' actions and the interpretation of those actions. Together, these factors affect the process of trusting, which in turn affect future agent actions, as the agents draw or apply information from the goal environment; their prior knowledge; interaction history; social norms; predispositions; and confidence. Two major publications have been submitted.
Bibliography Type:	Description: (Last Updated: 07/06/2025)