Toward Enabling Crew Resource Management Objectives in Human-Robot Collaboration in Safety-Critical Situations

Anushka Potdar
anushka.potdar@yale.edu
Yale University
New Haven, CT, USA

Tesca Fitzgerald
tesca.fitzgerald@yale.edu
Yale University
New Haven, CT, USA

ABSTRACT

Human-robot collaboration (HRC) has the potential to save human lives in safety-critical situations, such as robots collaborating with people in space exploration missions, search and rescue missions, firefighting, and disaster response. In order for robots to be able to assist humans in diagnosing and solving problems in these safety-critical situations, they need to be able to actively provide the social signals necessary to present high-dimensional data clearly, challenge a teammate's beliefs, collaboratively assess ambiguous problems, and design solutions. Crew Resource Management (CRM) offers a possible framework for how robots should exhibit these social signals. CRM is a proven strategy for facilitating effective human-human collaboration in safety-critical situations. In this paper, we analyze how current work in HRC and human-autonomy teaming aligns (and does not align) with the expectations set by CRM. We propose a framework that outlines the role of mental models in human-robot collaborative problem-solving, with the goal of supporting several key CRM objectives. Finally, we identify research questions that must be addressed to implement this framework, with the goal of enabling a robot to generate queries and responses that support the human-robot team's shared understanding of the problem.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative interaction**; • **Computing methodologies** → **Knowledge representation and reasoning**; **Cognitive robotics**.

KEYWORDS

human-robot collaboration, mental models, theory of mind, crew resource management

ACM Reference Format:
Anushka Potdar and Tesca Fitzgerald. 2024. Toward Enabling Crew Resource Management Objectives in Human-Robot Collaboration in Safety-Critical Situations. In *Proceedings of Workshop on Social Signal Modeling in Human-Robot Interaction (HRI '24).* ACM, New York, NY, USA, 6 pages.

1 INTRODUCTION

In NASA's Apollo 13 mission, when an oxygen tank explosion led to rapidly declining oxygen and electrical power [38], Mission Control Center supported the astronaut crew by monitoring and evaluating the results of the crew's actions, providing feedback, designing ideas, and sharing adapted plans [5]. This collaboration helped save the crew's lives, demonstrating the importance of decision-support systems in safety-critical situations. Yet, this support may not be available in long-duration exploration missions when communication with Mission Control Center is limited [40]. Similar needs arise in other safety-critical situations, such as aviation, firefighting, search and rescue, disaster response, and healthcare. How then can we develop robots that collaborate with humans to diagnose and solve problems in such situations?

In order for robots to be able to assist humans in diagnosing and solving problems in these safety-critical situations, they need to be able to actively provide the social signals necessary to present high-dimensional data clearly, challenge a teammate's beliefs, collaboratively assess ambiguous problems, and design solutions. Crew Resource Management (CRM) [17, 25–28, 30, 49, 55] offers a possible framework for how robots should exhibit these social signals. CRM is a proven strategy for facilitating effective human-human collaboration in safety-critical situations, providing guidelines for effective threat and error management, communication, decision making, team adaptability, and shared mental models [30].

In this paper, we analyze current work in human-robot collaboration (HRC) and human-autonomy teaming (HAT) with respect to CRM. We propose a framework that outlines the role of mental models in human-robot collaborative problem-solving, with the goal of supporting several key CRM objectives including threat and error management, standard operating procedures, communication, mutual performance monitoring, team leadership, decision making, team adaptability, and shared situational awareness.

Finally, we identify open research questions that must be addressed to implement this framework, with the goal of enabling a robot to generate queries and responses that support the human-robot team's shared understanding of the problem.

2 BACKGROUND

2.1 Crew Resource Management (CRM)

In 2009, U.S. Airways flight 1549 hit a flock of geese during take off, completely shutting down both engines [43]. Within only 3 minutes, Captain Sully successfully collaborated with the ground team and his co-pilot to safely land in the Hudson River and save all onboard. CRM helped make this possible [43].

CRM, initially developed to improve airline cockpit crews' teamwork, is the "application of human factors knowledge and skills to ensure that teams make effective use of all resources" [50]. "CRM encompasses effective and efficient error management through good communication, decision-making, feedback and conflict resolution,

HRI '24, March 11–15, 2024, Boulder, CO

workload management, and crew performance" [30]. Beyond aviation, CRM has also been successfully applied to other safety-critical domains including spaceflight, healthcare, maritime, gas, and rail industries [30].

CRM's core training concepts include: threat and error management, verbalize verify monitor, standard operating procedures, communication, briefing, backup behavior, mutual performance monitoring, team leadership, decision making, task-related assertiveness, team adaptability, and shared situational awareness [30]. Shared situational awareness, also known as *shared mental models* (SMMs), refers to team members' ability to "gather and use information to develop a common understanding of the task and team environment" [30]. SMMs have been identified as the most robust signal of effective teaming [9, 10, 34] and an indicator of creative problem-solving in analog crews experiencing prolonged isolation and confinement [14]. This is particularly important in situations of acute stress and anxiety [6], where an individual's attention and cognitive tunneling "could result in failure to detect potentially critical events" [29].

2.2 Human Autonomy Teaming (HAT)

HAT "involves humans working interdependently toward a common goal along with autonomous agents" [46]. Shively et. al. identified that HAT research has a lot to learn from CRM and that it is beneficial to expand CRM to include collaboration between automated teammates and humans [50]. They also identified that (1) the existing HAT concepts bi-directional communication and working agreements have the potential to support CRM-like behavior and (2) using these existing HAT concepts, automation can be designed to mirror CRM skills [50]:

- **Bi-directional communication**: This enables "humans to team effectively with automation and allows the human (and automation) to question, share hypotheses, provide additional input, etc. just as human teammates would" [50].
- **Working agreements**: In CRM, standard operating procedures include checklists and standardized callouts and aim to "ensure that a known, safe, efficient set of actions is used to navigate through complex procedures that require great accuracy" [50]. HAT has a similar concept to standard operating procedures [21, 42] "working agreements, which encapsulate goals, procedures, and division of responsibility into a package that can be specified offline and instantiated quickly in real-time situations" [50].

2.3 Mental Models

Mental models have been defined as "organized knowledge structures that [...] help people to describe, explain, and predict events in their environment" [41]. There are different types of mental models including ones that model procedures for tasks, team member's preferences and abilities, and interaction between team members [41][54].

Shared Mental Models. Decades of organizational psychology research showing that SMMs improve teamwork in humans [15, 19, 37, 52] has led to work on SMMs in HRI [19, 45]. Gervits et. al. were the first to introduce a computational framework for both robot-robot SMMs (SMMs *between robots*) and human-robot SMMs [19]. They were also the first to test the hypothesis that SMMs improve human-agent teams' coordination and performance, showing that robot-robot SMMs improve overall human-robot team performance [19]. Nikolaidis and Shah worked on human-robot cross-training, where a robot and human learn a shared collaborative task plan by alternating roles [45], focused on the human and robot's task execution and enabling a robot to adapt to a person's established workflow patterns. They showed that emulating proven human teamwork methods, like cross-training, may be the best way to achieve "effective and fluent human-robot teaming" [45].

Toward Theory of Mind (ToM). Being able to predict the mental states of others is critical for distributed multi-agent systems that need to communicate and cooperate. In human-human interactions, ToM is key to communicating information in conversations [36] and in maintaining consensus during collaboration and communication. [3]. ToM "is the ability to explain, predict, and interpret behavior by attributing mental states such as desires, beliefs, intentions and emotions to oneself and to other people" [13]. When operating as a part of a team, these inferences are necessary for deciding *when* and *with whom* one should share its own intentions in order to reach consensus within the team [58]. Humans naturally infer mental models of their teammates' beliefs and goals based on their teammates' dialogue [7]. Enabling ToM between humans and robots is a challenge, and often inferred models do not reflect the truth, which can be framed as a model reconciliation problem [11]. Although there has been a recent rise in modeling multi-agent collaboration [12, 18, 33, 53], integrating ToM remains a significant challenge [60]. Mental models can not be directly observed, but we can infer them using observable evidence including dialogue. [2].

3 PROBLEM FORMULATION

Prior work on HRC has generally focused on human-robot *coordination* [32, 44] where the human and robot *work independently* in a shared space toward a shared goal. We focus on a human and robot *working together* to diagnose and solve complex problems, where both the robot and human take initiative to piece together their partial information toward a solution.

We represent the human-robot SMM as a combination of (1) what the robot believes about the problem and (2) what the robot believes the human believes about the problem. Consider this example: onboard a deep space vehicle, during lack of communication with Mission Control Center, a time-critical problem occurs within the Environmental Control and Life Support System, and a robot collaborates with the astronaut crew team to diagnose and solve the problem.

Here, the SMM would represent (1) the robot's belief that a valve is working properly, *and* (2) the human's belief that the valve is not working properly. It is possible for this SMM to reflect *conflicting* beliefs held by the human and robot. Alternatively, a *consensus* SMM would only represent the beliefs the human and robot agree on. We view consensus building as a step after building a SMM that involves managing conflicts to build human-robot agreement, which we plan to address in our future work. Our goal is for the robot to generate queries and responses that are *jointly* informative for both the robot and human's understanding of the problem so that they can identify an appropriate solution.

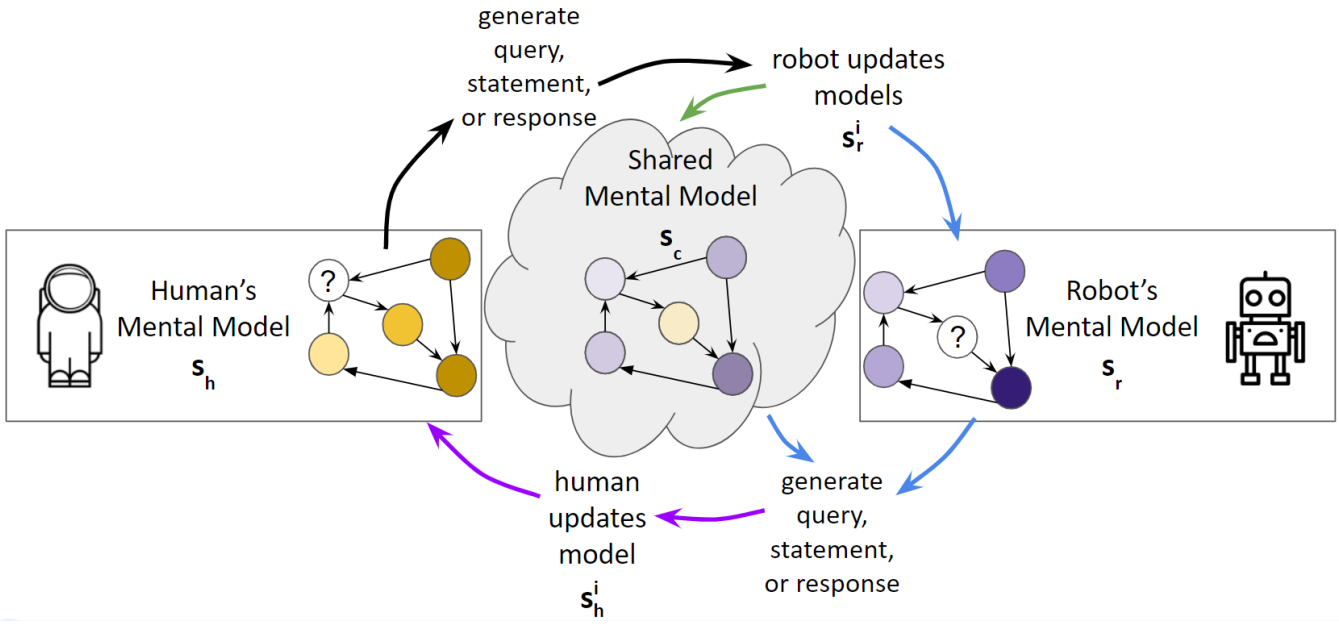

Figure 1: In the human's (s_h) and the robot's (s_r) knowledge graph representation of their mental model, each node indicates their belief and confidence over the current status of a particular spacecraft system's component. We propose using interaction between the robot and the crew to create a shared mental model s_c.

3.1 Framework for Human-Robot Collaborative Problem-Solving

In Fig. 1, we represent the problem the human-robot team aims to solve as a knowledge graph. Referencing our previous example, this knowledge graph contains one node for each component of the Environmental Control and Life Support System (i.e., oxygen pressure gauge, each sensor) and edges between nodes to indicate how they influence each other. The graph's structure is initialized to reflect the extensive training astronauts receive, including information in the standard operating procedures (i.e., crew checklists). There are three total such knowledge graphs that each represent the mental model *of an agent's belief over the problem state*: one for the human's, one for the robot's, and one for the human-robot SMM.

The specific *true system state* of the knowledge graphs' nodes (i.e., the state of each component in the system) and edges (i.e., how the components influence each other) is unknown to the robot and human. During the problem-solving process, the team's understanding of how the system's components influence one another can change (especially if the system is in an anomaly state where these relationships do not match the crew's training knowledge, including information in the standard operating procedures about the system).

The human-robot team may even have an incorrect understanding of the graph structure. For example, in a spacewalk mishap (U.S. Extravehicular Activity 23), the crew and International Space Station community attempted to diagnose the source of water intrusion into an astronaut's helmet. Despite their expertise on spacesuit systems, they mistakenly attributed the problem to a leaking drinking bag [1], when in fact it was caused by "a clog inside the EMU [spacesuit] Fan Pump Separator, caused by inorganic material that led to water spilling into the vent loop" [24]. If we consider our framework's knowledge graph mental model representation of the crew in this example, one of the inaccuracies of the knowledge graph structure is that it did not have an accurate edge connection between the vent loop node and the helmet node.

The goal of the human-robot team is to collaborate in order to identify the true, hidden state of the system. We formalize this collaboration as the following parts:

- The **robot's mental model $\mathbf{s_r}$** contains the robot's belief of the system state and is informed by the robot's sensor readings. This model may be incomplete (due to a lack of sensors for some components) or incorrect (due to faulty sensor readings).
- The **human's mental model $\mathbf{s_h}$** contains the human's belief of the system state and is informed by the human's access to sensor readings, physical observations of the system, and the human's training knowledge of the system. This model may also be incomplete or incorrect.
- The **shared mental model $\mathbf{s_c}$** contains the robot's estimate of the robot and crew's hidden shared mental state, which is built over time through the human and robot generating queries and responses that are *jointly* informative for both the robot and human's understanding of a problem so that they can collaboratively identify an appropriate solution.

The robot can generate queries in order to learn from the human's understanding of the problem state and to suggest diagnostic measures. In doing so, however, the robot's queries are likely to also influence the human's understanding of the problem [22, 35, 47, 51].

The robot should leverage this phenomenon in order to ask questions that are not only informative to its own model, but also toward a SMM of the problem state $\mathbf{s_c}$ with the crew because studies show that having a SMM is key to effective teaming [9, 10, 34]. As shown in Fig. 1, the SMM is built over time through each query-response interaction. This essentially involves performing a belief update $P(S_c^{i+1}|S_c^i, \langle q, r\rangle)$ for a Hidden Markov Model (HMM), where the crew response r to a query q is an emission of the hidden SMM s_c at query-response iteration i.

Working toward a consensus of the problem state is not always optimal. Prior work has identified several key factors for successful and creative problem-solving [6, 39]. Of these key factors is managing openness which requires revisiting old beliefs about the problem state that may be incorrect and prevent the team from being able to reconcile new evidence [59]. Problem-solving requires a synergy between working toward a shared consensus and challenging beliefs. In our future work, we aim to emulate key practices in human-human *creative* problem-solving in HRC.

4 OPEN RESEARCH QUESTIONS

Based on this framework, **our primary goal is for the robot to produce queries and responses that are maximally informative to the human-robot team's shared understanding of the problem**. Toward this objective, we identify two open research questions:

- **RQ1**: How can we infer the human's mental model through their dialogue (query or response) with the robot?
- **RQ2**: How does a robot's dialogue influence the human's mental model of the problem?

4.1 Inferring Mental Models Through Dialogue

We aim to maximize information gain with respect to the human and robot's shared understanding of the problem. A key challenge to accomplishing this aim is this: How will the robot estimate this shared understanding in our framework? The green arrow in Fig. 1 represents this challenge. To estimate this shared understanding, the robot must (1) estimate the human's mental model and (2) compare the human's mental model and its own mental model to estimate the SMM. This challenge requires us to investigate **RQ1**.

Proposed Work. We propose adapting and extending Briggs and Scheutz's belief update rules for utterances [7] to our problem. We will extend them to dynamically update the knowledge graph representations of the human, robot, and shared mental models. More specifically, as shown in Fig. 1, we aim to build a SMM over time through multiple query-response interactions. We expect that a key challenge will involve adapting the belief update rules so they are compatible with our knowledge graph representations and then extending these rules to support our complex collaborative problem-solving dialogue context. Briggs and Scheutz's belief update rules are simplified to provide a starting point. A specific example of their simplification is that their belief update rules assume that "an agent always believes all propositions it is able to infer from the utterance of another agent" [7]. We will need to modify this for our problem statement because collaborative problem-solving requires these belief update rules to support a variety of complex behaviors including challenging beliefs and reconsidering past beliefs.

4.2 Predicting How Dialogue Influences Mental Models

The previous RQ involved estimating the human's mental model based on their utterances, but what influences that mental model in the first place? Prior work shows that a robot's dialogue can influence a human's perception of the robot's abilities [16, 23]. It can also build a human's trust in the robot's decision-making (including inappropriate trust [31, 48]). In order to maximize *accurate* shared understanding of the problem, it is important to model how the robot's own dialogue is likely to affect the human's mental model. The blue arrows in Fig. 1 represent this problem of generating queries and responses, with the goal of creating a SMM of the problem state. In order to achieve this goal, it is essential that we understand how the robot's dialogue influences the human's mental model (and ultimately, the SMM).

Humans' ability to predict others' future actions is key to successful social interactions [56]. Beun [4] views the purpose of dialogue as being to influence "the relevant aspects of the mental state of a recipient," and one of the first models of mental states [20] defined the purpose of dialogue as being "to change the interlocutor's mental state and reach the goals of the interaction" [8]. If the purpose of dialogue involves changing the recipient's mental model in a specific way towards a goal, then it is essential that the speaker can predict their dialogue's impact on the recipient. This leads us to **RQ2**.

Most work related to this RQ centers on the role of trust in HRI [48][57] and robots communicating their abilities [23], but does not offer other explanations for how the robot's dialogue may influence the human's mental model of the problem.

Proposed Work. This requires us to introduce a new model that predicts how a robot's queries and responses influence the human's mental model of the problem. This is an open problem that could be addressed by a user study that evaluates how people infer knowledge from a robot's dialogue.

5 CONCLUSION

In order for robots to be able to assist humans in diagnosing and solving problems in safety-critical situations, they need to be able to actively provide the social signals necessary to present high-dimensional data clearly, challenge a teammate's beliefs, collaboratively assess ambiguous problems, and design solutions. In this paper, we analyzed how current work in HRC and HAT aligns (and does not align) with the expectations set by CRM. We then presented a framework for addressing this problem via the lens of CRM. Finally, we identified two key research questions to guide our future work on implementing this framework, with the goal of enabling a robot to generate queries and responses that support the human-robot team's shared understanding of the problem.

REFERENCES

[1] 2014. Mishap Investigation Board Briefing on Spacesuit Water Intrusion APPEL Knowledge Services. https://appel.nasa.gov/2014/02/27/mishap-investigation-board-briefing-on-spacesuit-water-intrusion/

[2] Francesca Alloatti, Federica Cena, Luigi Di Caro, Roger Ferrod, Giovanni Siragusa, et al. 2021. Towards Mental Model-driven Conversations. In *CEUR WORKSHOP PROCEEDINGS*, Vol. 2903. CEUR-WS, 1–5.

[3] Cristian-Paul Bara, Sky CH-Wang, and Joyce Chai. 2021. MindCraft: Theory of mind modeling for situated dialogue in collaborative tasks. *arXiv preprint arXiv:2109.06275* (2021).

[4] Robbert-Jan Beun. 1994. Mental state recognition and communicative effects. *Journal of Pragmatics* 21, 2 (1994), 191–214.

[5] Samira Bourgeois-Bougrine. 2020. What does creativity mean in safety-critical environments? *Frontiers in Psychology* 11 (2020), 565884.

[6] s Bourgeois-Bougrine. 2020. What Does Creativity Mean in Safety-Critical Environments? *Frontiers in Psychology* 11 (10 2020). https://doi.org/10.3389/fpsyg.2020.565884

[7] Gordon Briggs and Matthias Scheutz. 2012. Multi-modal belief updates in multi-robot human-robot dialogue interactions. *AISB/IACAP World Congress 2012: Linguistic and Cognitive Approaches to Dialogue Agents, Part of Alan Turing Year 2012* (Jan. 2012), 67–72.

[8] Zoraida Callejas, David Griol, and Ramón López-Cózar. 2011. Predicting user mental states in spoken dialogue systems. *EURASIP Journal on Advances in Signal Processing* 2011, 1 (2011), 1–21.

[9] Janis A Cannon-Bowers, Eduardo Salas, and SA Converse. 1990. Cognitive psychology and team training: Training shared mental models and complex systems. *Human factors society bulletin* 33, 12 (1990), 1–4.

[10] Janis A Cannon-Bowers, Eduardo Salas, and Sharolyn Converse. 1993. Shared mental models in expert team decision making. *Individual and group decision making: Current issues* (1993).

[11] Tathagata Chakraborti, Sarath Sreedharan, Yu Zhang, and Subbarao Kambhampati. 2017. Plan explanations as model reconciliation: Moving beyond explanation as soliloquy. *arXiv preprint arXiv:1701.08317* (2017).

[12] Abhishek Das, Satwik Kottur, José MF Moura, Stefan Lee, and Dhruv Batra. 2017. Learning cooperative visual dialog agents with deep reinforcement learning. In *Proceedings of the IEEE international conference on computer vision*. 2951–2960.

[13] Jean Decety and Margarita Svetlova. 2012. Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental cognitive neuroscience* 2, 1 (2012), 1–24.

[14] Leslie A. DeChurch, Alina Lungeanu, and Noshir S. Contractor. 2023. Think like a team: Shared mental models predict creativity and problem-solving in space analogs. *Acta Astronautica* (2023). https://doi.org/10.1016/j.actaastro.2023.10.022

[15] Alberto Espinosa, Robert Kraut, Sandra Slaughter, Javier Lerch, James Herbsleb, and Audris Mockus. 2002. Shared mental models, familiarity, and coordination: A multi-method study of distributed software teams. *ICIS 2002 Proceedings* (2002), 39.

[16] Kerstin Fischer. 2018. When transparent does not mean explainable. In *Workshop on Explainable Robotic Systems*.

[17] Rhona Flin and Lynne Martin. 2001. Behavioral markers for crew resource management: A review of current practice. *The International Journal of Aviation Psychology* 11, 1 (2001), 95–118.

[18] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. 2018. Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[19] Felix Gervits, Dean Thurston, Ravenna Thielstrom, Terry Fong, Quinn Pham, and Matthias Scheutz. 2020. Toward Genuine Robot Teammates: Improving Human-Robot Team Performance Using Robot Shared Mental Models.. In *Aamas*. 429–437.

[20] Jonathan Ginzburg et al. 1996. Dynamics and the semantics of dialogue. *Logic, language and computation* 1 (1996), 221–237.

[21] Robert S Gutzwiller, Sarah H Espinosa, Caitlin Kenny, and Douglas S Lange. 2018. A design pattern for working agreements in human-autonomy teaming. In *Advances in Human Factors in Simulation and Modeling: Proceedings of the AHFE 2017 International Conference on Human Factors in Simulation and Modeling, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8*. Springer, 12–24.

[22] Soheil Habibian, Ananth Jonnavittula, and Dylan P. Losey. 2022. Here's What I've Learned: Asking Questions That Reveal Reward Learning. *J. Hum.-Robot Interact.* 11, 4, Article 40 (sep 2022), 28 pages. https://doi.org/10.1145/3526107

[23] Soheil Habibian, Ananth Jonnavittula, and Dylan P. Losey. 2022. Here's What I've Learned: Asking Questions That Reveal Reward Learning. *J. Hum.-Robot Interact.* 11, 4, Article 40 (sep 2022), 28 pages. https://doi.org/10.1145/3526107

[24] Christopher Hansen and Christopher Cassidy. 2014. Mishap Investigation Board Summary of Extravehicular Activity 23: Lessons Learned From a Spacewalk Close Call. *Journal of Space Safety Engineering* 1, 1 (2014), 32–39.

[25] Robert L Helmreich. 1991. The long and short term impact of crew resource management training. In *Proceedings of the AIAA, NASA, FAA, and Human Factors Society Conference on Challenges in Aviation Human Factors: The National Plan*. 81–83.

[26] Robert L Helmreich and Ashleigh C Merritt. 2017. 11 Safety and error management: The role of crew resource management. In *Aviation Resource Management: Proceedings of the Fourth Australian Aviation Psychology Symposium Volume 1*. Routledge.

[27] Robert L Helmreich, Ashleigh C Merritt, and John A Wilhelm. 2017. The evolution of crew resource management training in commercial aviation. In *Human error in aviation*. Routledge, 275–288.

[28] Robert L Helmreich and John A Wilhelm. 1991. Outcomes of crew resource management training. *The International journal of aviation psychology* 1, 4 (1991), 287–300.

[29] Jerzy Jarmasz, Chris Herdman, and Kamilla Johannsdottir. 2005. Object-Based Attention and Cognitive Tunneling. *Journal of experimental psychology. Applied* 11 (03 2005), 3–12. https://doi.org/10.1037/1076-898X.11.1.3

[30] Barbara G Kanki, José Anca, and Thomas R Chidester. 2019. *Crew resource management*. Academic Press.

[31] Ulas Berk Karli, Shiye Cao, and Chien-Ming Huang. 2023. " What If It Is Wrong": Effects of Power Dynamics and Trust Repair Strategy on Trust and Compliance in HRI. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 271–280.

[32] Maram Khatib, Khaled Al Khudir, and Alessandro De Luca. 2017. Visual coordination task for human-robot collaboration. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 3762–3768.

[33] Michael Kinney and Costas Tsatsoulis. 1998. Learning communication strategies in multiagent systems. *Applied intelligence* 9 (1998), 71–91.

[34] Richard Klimoski and Susan Mohammed. 1994. Team mental model: Construct or metaphor? *Journal of management* 20, 2 (1994), 403–437.

[35] Sachin G Konan, Esmaeil Seraj, and Matthew Gombolay. 2022. Iterated Reasoning with Mutual Information in Cooperative and Byzantine Decentralized Teaming. In *International Conference on Learning Representations*. https://openreview.net/forum?id=giBFoa-uS12

[36] Meredyth Krych-Appelbaum, Julie Banzon Law, Dayna Jones, Allyson Barnacz, Amanda Johnson, and Julian Paul Keenan. 2007. "I think I know what you mean": The role of theory of mind in collaborative communication. *Interaction Studies* 8, 2 (2007), 267–280.

[37] M Lee, Tristan Johnson, and Myung H Jin. 2012. Toward understanding the dynamic relationship between team and task shared mental models as determinants of team and individual performances. *International journal of information technology and business management* 8, 1 (2012), 1–14.

[38] Jim Lovell and Jeffrey Kluger. 2006. *Apollo 13*. Houghton Mifflin Harcourt.

[39] Todd Lubart, Franck Zenasni, and Baptiste Barbot. 2013. Creative Potential and Its Measurement. *International Journal for Talent Development and Creativity* 1 (12 2013), 41–50.

[40] Jessica J. Marquez, Steven Hillenius, Bob Kanefsky, Jimin Zheng, Ivonne Deliz, and Marcum Reagan. 2017. Increasing crew autonomy for long duration exploration missions: Self-scheduling. In *2017 IEEE Aerospace Conference*. IEEE, Big Sky, MT, USA, 1–10. https://doi.org/10.1109/AERO.2017.7943838

[41] John E Mathieu, Tonia S Heffner, Gerald F Goodwin, Eduardo Salas, and Janis A Cannon-Bowers. 2000. The influence of shared mental models on team process and performance. *Journal of applied psychology* 85, 2 (2000), 273.

[42] Christopher A Miller and Raja Parasuraman. 2007. Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. *Human factors* 49, 1 (2007), 57–75.

[43] Jerry Mulenburg. 2011. Crew Resource Management Improves Decision Making | APPEL Knowledge Services. https://appel.nasa.gov/2011/05/11/crew-resource-management-improves-decision-making/

[44] Bilge Mutlu, Allison Terrell, and Chien-Ming Huang. 2013. Coordination mechanisms in human-robot collaboration. In *Proceedings of the Workshop on Collaborative Manipulation, 8th ACM/IEEE International Conference on Human-Robot Interaction*. Citeseer, 1–6.

[45] Stefanos Nikolaidis and Julie Shah. 2013. Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy. In *2013 8th ACM/IEEE international conference on human-robot interaction (HRI)*. IEEE, 33–40.

[46] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. 2022. Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors* 64, 5 (2022), 904–938.

[47] Rohan Paleja, Muyleng Ghuy, Nadun Ranawaka Arachchige, Reed Jensen, and Matthew Gombolay. 2021. The utility of explainable ai in ad hoc human-machine teaming. *Advances in neural information processing systems* 34 (2021), 610–623.

[48] Paul Robinette, Ayanna M Howard, and Alan R Wagner. 2015. Timing is key for robot trust repair. In *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*. Springer, 574–583.

[49] Eduardo Salas, Jennifer E Fowlkes, Renee J Stout, Dana M Milanovich, and Carolyn Prince. 1999. Does CRM training improve teamwork skills in the cockpit?: Two evaluation studies. *Human Factors* 41, 2 (1999), 326–343.

[50] Robert J Shively, Joel Lachter, Robert Koteskey, and Summer L Brandt. 2018. Crew resource management for automated teammates (CRM-A). In *Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings 15*. Springer, 215–229.

[51] Andrew Silva, Mariah Schrum, Erin Hedlund-Botti, Nakul Gopalan, and Matthew Gombolay. 2023. Explainable Artificial Intelligence: Evaluating the Objective and Subjective Impacts of xAI on Human-Agent Interaction. *International Journal of Human–Computer Interaction* 39, 7 (2023), 1390–1404. https://doi.org/10.1080/

10447318.2022.2101698 arXiv:https://doi.org/10.1080/10447318.2022.2101698

[52] Renée J Stout, Janis A Cannon-Bowers, Eduardo Salas, and Dana M Milanovich. 1999. Planning, shared mental models, and coordinated performance: An empirical link is established. *Human factors* 41, 1 (1999), 61–71.

[53] Sainbayar Sukhbaatar, Rob Fergus, et al. 2016. Learning multiagent communication with backpropagation. *Advances in neural information processing systems* 29 (2016).

[54] Aaquib Tabrez, Matthew B Luebbers, and Bradley Hayes. 2020. A survey of mental modeling techniques in human–robot teaming. *Current Robotics Reports* 1 (2020), 259–267.

[55] JC Taylor and MM Robertson. 1995. *The effects of Crew Resource Management (CRM) training in airline maintenance: Results following three year's experience.* Technical Report.

[56] Mark A Thornton and Diana I Tamir. 2017. Mental models accurately predict emotion transitions. *Proceedings of the National Academy of Sciences* 114, 23 (2017), 5982–5987.

[57] Alan R Wagner and Paul Robinette. 2021. An explanation is not an excuse: Trust calibration in an age of transparent robots. In *Trust in Human-Robot Interaction.* Elsevier, 197–208.

[58] Yuanfei Wang, Fangwei Zhong, Jing Xu, and Yizhou Wang. 2022. ToM2C: Target-oriented Multi-agent Communication and Cooperation with Theory of Mind. In *International Conference on Learning Representations.* https://openreview.net/forum?id=M3tw78MH1Bk

[59] Jennifer Watts-Perotti and David D Woods. 2009. Cooperative advocacy: an approach for integrating diverse perspectives in anomaly response. *Computer Supported Cooperative Work (CSCW)* 18 (2009), 175–198.

[60] Luyao Yuan, Zipeng Fu, Linqi Zhou, Kexin Yang, and Song-Chun Zhu. 2021. Emergence of theory of mind collaboration in multiagent systems. *arXiv preprint arXiv:2110.00121* (2021).

Received 16 February 2024; revised 12 March 2009; accepted 5 June 2009