Human Factors and Simulation, Vol. 83, 2023, 46–56
https://doi.org/10.54941/ahfe1003561

The Role of Artificial Theory of Mind in Supporting Human-Agent Teaming Interactions

Jessica Williams[1], Rhyse Bendell[1], Stephen M. Fiore[2], and Florian Jentsch[1]

[1]Team Performance Laboratory, Institute for Simulation and Training, University of Central Florida, Orlando, FL 32826, United States
[2]Cognitive Sciences Laboratory, Institute for Simulation and Training, University of Central Florida, Orlando, FL 32826, United States

ABSTRACT

In this article we discuss the role of Artificial Theory of Mind (AToM) in supporting human-agent teaming interactions. Humans are able to interpret, understand, and predict another's behavior by leveraging core socio-cognitive processes, generally referred to as Theory of Mind (ToM). A human's ToM is critical to their ability to successfully interact with others, especially in the context of teaming. Considering the increasing role of AI in team cognition, there is an emerging need for agents capable of such complex socio-cognitive processes. We report findings from a large multi-organization research program, DARPA's Artificial Social Intelligence Supporting Teams (ASIST), designed to study teamwork with socially intelligent artificial agents serving as team advisors. We focus on agent-to-human communications, including content, intended purpose, and, particularly, the use of AToM attributions in both covert agent explanations as rationale for giving a certain intervention, as well as the use of agents making overt ToM attributions of players in the intervention itself. The findings suggest that agent teammates are able to demonstrate AToM and that that interventions based upon these can influence team outcomes. We discuss the impact of the various types of ASI interventions and their effect on teams, and provide recommendations for future research on human-AI teaming.

Keywords: Artificial social intelligence, Artificial theory of mind, Human-agent teams

INTRODUCTION

In our day-to-day interactions, humans are able to interpret, understand, and predict another's behavior by leveraging a core set of socio-cognitive process. Generally referred to as Theory of Mind (ToM), this enables humans to effectively interact with others (Cuzzolin et al., 2020). Due to the increasing sophistication of artificial intelligence (AI), agents are taking on more collaborative roles in the modern workplace. Although team research is currently studying how AI as a teammate influences teamwork, the majority of this research focuses on AI capable of executing task relevant objectives (McNeese et al., 2018; Demir et al., 2020). In order for AI to take on more collaborative responsibilities, it must also have socio-cognitive competence.

This necessitates AI capable of perceiving, interpreting, and extracting the social information contained in interactions with others (Alonso & De La Puente, 2018). This requires AI capable of understanding and simulating human mental states, such as their beliefs, intentions, emotions, and motivations, what we refer to as Artificial Theory of Mind (AToM; Williams et al., 2022). AToM is critical for agents that interact with humans, especially in teams or multi-agent systems (Oguntola et al., 2021).

Research in human-human teams shows that high-performing teams are able to coordinate their specialized roles to provide task-relevant knowledge that anticipates each other's needs and supports each other's taskwork to engage in complex problem-solving and teaming behaviors (Cooke et al., 2012; DeChurch & Mesmer-Magnus, 2010; Fiore & Schooler, 2004; Newton & Fiore, 2020). This additionally necessitates artificial social intelligence and theory of mind in order to build shared mental models of their team and task (Fiore et al., 2021), that is, the shared knowledge and awareness about how members understand and will respond in operational settings (Mathieu et al., 2000). Likewise, for an ASI agent to be able to anticipate human needs to better predict and respond to human requests in a way that accounts for their context (e.g., emotional state or taskwork) it will need to have developed an AToM (Williams et al., 2022). An agent utilizing AToM would be afforded the ability to recognize incorrect or incomplete beliefs and knowledge, which provide opportunities for correction through the use of appropriately-timed and contextually situated interventions (Oguntola et al., 2021).

Research on human-machine interaction and social cognition currently focuses on what is called social signals processing (SSP; (Fiore et al., 2013; Vinciarelli et al., 2009). Here, sensors, such as video and audio, are used to capture movement and sound dynamics used to infer mental states of a human interactor (Wiltshire et al., 2014) directly, or through analysis of extractions such as natural language processing (Basavaraj et al., 2022). Despite substantial progress in this area, SSP research is mostly studied using simple interactions and confined to carefully controlled environments. Further, technology capable of capturing and interpreting the rich array of social cues emerging from a blend of context, non-verbal behavior, dynamic interactions, and speech, is still limited. This, then, limits the ability to understand how agents capable of AToM influence teamwork.

To compensate for these limitations, the current research studied AToM and team interaction using a tightly designed evaluative framework: a simulated Minecraft-based testbed for human-agent teamwork (Corral et al., 2021; Freeman et al., 2023) simulating a search and rescue team task (Cooke & Shope, 2017). The tasks mimicked constraints that a typical USAR team may experience through implementing a dynamic environment with in-game perturbations and time-sensitive tasks that required coordination. Team interactions and channels for social signals were codified through a message bus, allowing agents access to cues providing the foundation for AToM (Freeman et al., 2023). Because the ability to perceive, interpret, and understand is critical to ToM, using a simulated testbed provided AI with the foundational data they could use to develop AToM. In this context, then, it is possible to study ASI in human-agent teaming and better understand how this affects team processes and outcomes.

We next report a portion of the findings from a large multi-university experiment studying ASI in human-agent teams. In this experiment, ASI acted as coaches or facilitators of teamwork. The description of AI architecture is beyond the scope of this paper (see Freeman et al. 2023; Huang et al. 2022b). We report on the outcomes of the ASI facilitation. Specifically, ASI monitored team interactions, objectives met, points, earned, etc. and generated interventions devised to improve teamwork. These interventions were delivered as brief prompts set to either the entire team, or specific team members. We focus on the prompts and report our analyses of their content as they relate to teaming and AToM.

METHOD

Overview of the Experimental Task and Agent Capabilities

The data for our analyses was collected as part of an experiment run by Arizona State University as part of the third program experiment of the "Artificial Social Intelligence for Supporting Teams" (ASIST) program run by the Defense Advanced Projects Research Agency (DARPA) (see Huang et al., 2022a). The data analyzed in this paper is available in the dataset (available here: Huang et al., 2022b). The study used a Minecraft-based testbed for a game-based simulated Urban Search and Rescue (USAR) experimental task (Huang et al., 2022a). The SAR missions were constructed such that team members had to gather and integrate information and coordinate their actions to meet objectives. The different ASI enabled agents were individually used in their own missions where they acted as advisors for the team. Additionally, for a portion of the missions, a human advisor was used to act as a comparison to the ASI. The reader is referred to these for detail on the overall experiment and the measures used. Here we focus only on those parts of the experiment that were related to the agent prompts that were the basis for the only interactions between the AI and human team members.

First, prior to the experiment, participants were given a set of surveys capturing individual differences in areas potentially predictive of differences in teamwork and taskwork. This included a video game experience measure (Bendell et al., 2020), the Reading the Mind in the Eyes Task (RMET; Baron-Cohen et al., 2001) and Santa Barbara Sense of Direction (SSOD; Hegarty et al., 2002). Based upon a theoretical framework of team and task competencies (Bendell et al., 2021), results of these surveys were used to create individual and team profiles predicting potential performance in the testbed. Data from these surveys were used by a portion of the agents to inform the models and the particular mental state attributions leading to ASI based interventions. Participants also received training on the testbed mechanics, mission objectives, and rules and tools available. A key tool developed to study coordination was based upon extended cognition theory and the use of artifacts to offload team cognition into the game environment (Fiore & Wiltshire, 2016). These were therefore used as means of communicating to teammates task relevant data pertinent to mission objectives. Each participant had the same set of marker blocks they could utilize to externalize their communication; the semantic meaning of each marker block available was as follows: Regular victim here, Critical victim here, No victim here, Victim injury type A, Victim injury type B, Victim injury type C, Rubble here, Threat room here, and help me here. Thus, the marker blocks afforded participants the opportunity to externalize their communication via shared cognitive artifacts. Although the six ASI agents (e.g., Li et al., 2023; Wang et al., 2023) were designed individually by six separate ASIST program teams all agents were able to track the team members in the mission, what actions they took in the environment, and what marker blocks they placed, and where they placed them. For more detail on the ASI enabled agents, the reader is referred to Huang et al. (2022a) and Freeman et al. (2023). Thus, all agents provided the same kind of information that the human advisor was able to acquire by monitoring teams during their missions (i.e., agents use of the message bus was equivalent to the human advisor viewing a video feed (Freeman et al., 2023). However, neither the human advisor or agents had access to the 'ground truth' information channels that contained mission relevant information (e.g., knowledge integration task clues, the locations of victims and threat rooms (Huang et al., 2022a). The ASI used this real-time trial data in forming what we are calling their AToM, which was then used to inform the interventions they provided to the team members individually, as a dyad, and as a whole.

Intervention Coding

The various interventions across all ASI advisor trials were extracted from the trial messages files (for this data, see: Huang et al., 2022b) and then compiled into one datasheet for all teams in the ASI advisor condition group, regardless of which ASI advisor they were assigned, so that we are able to look across the different agents and at the features of the intervention prompts as a whole. We began by filtering the list of interventions to identify and remove prompts that were exact duplicates of each other, so that the remaining list of interventions would be comprised of all, and only, the unique interventions. This meant that some interventions were nearly identical, some only differing by specific numerical values. Thus, the interventions that only differed by a number (e.g., "there are [n] non-critical victims and [n] critical victims stabilized") or specific player reference (e.g., referring to Medic, Transporter, Engineer), and not by the content of the intervention, would receive the same code. For example, the intervention "1 non-critical and 1 critical victims stabilized. 2 victims transported. GO TEAM!" was given the primary code 'Mission Status Update' and the secondary code 'Motivation' (intervention categories discussed below), and the intervention "9 non-critical and 4 critical victims stabilized. 12 victims transported. GO TEAM!" received the same codes.

To determine the categories for coding the interventions, the first three authors collaboratively conducted an initial reflexive thematic analysis. This involved generation of an initial 1st pass at categories to label the interventions. From this, we separately reviewed the categories and unique interventions list, making a note of any new categories we found appropriate to add to the list (note, only one new category was added). The coders then discussed the categories, further refined them, generated examples, and we reached agreement on all categories. Once they unanimously agreed on category definitions, the three coders applied first, a primary code to each intervention and, where necessary, a secondary code. Not every intervention required a secondary code, but these secondary codes were included when the coders agreed it was necessary to indicate a greater depth of detail. Once every unique intervention had been assigned a primary code and, if necessary, secondary code, those codes were then carried through to all instances of the ASI interventions. The full list of the code categories was agreed upon, and a description providing exemplars of cues or statements from the interventions that help to scope the category's application (see Table 1.). Additionally, we have provided the percentages of each category's occurrence. The first three authors individually coded the interventions then discussed the coding, and any differences were discussed until consensus was met.

Below, we provide examples of interventions that were given by the ASI agents and the primary and secondary codes that were assigned to them (see Table 2.). Relevant to AToM, a subset of the interventions draws upon theory of mind by explicitly making attributions. This included first-order

Table 1. Coding category list.

Category	Category Description	% of All Interventions Given
Advisor Role Capability	I will assist you in this mission, I will be providing advice, I will alert you	1.58%
Coordination	sync up with…., work closely with	8.40%
Explicit Communication	remind your teammates, talk to…, ask for help	0.36%
Explicit Communication: Marker	let your teammate know about the marker, remind team to place markers	9.98%
Explicit Communication: Strategy	should talk to…, ask the Engineer	1.41%
Explicit Communication: ToM	it seems you…, who appears to be…, may be confused…	14.45%
Implicit Communication: Marker	you missed marking, keep placing markers, use the … marker	9.03%
Information Sharing	…can help the team by telling about the C8 shortcut, there are … stabilized victims	3.22%
Interpersonal Process	Hello from the agent	1.61%
Mission Status Update	you have transported N victims,... critical and... noncritical victims	6.50%
Motivation	Bravo!, Great job, congratulations, keep going	30.20%
Strategy General	prioritize, rethink your coordination, can anyone help,	6.24%
Strategy Role Capability	apply more role-specific skills, let … transport since they are faster,	3.97%
Strategy Sequence	after, start evacuating, next	3.05%

Table 2. Intervention coding examples.

Intervention	Primary Code	Secondary Code
Blue, it seems you need some help to rescue a critical victim. Ask your teammates for assistance.	Explicit Communication: ToM	Explicit Communication: Strategy
Be sure to use the room marker to decide whether to enter this room.	Implicit Communication: Marker	Strategy General
When evacuating a victim, utilize your map to find the closer evacuation zone next time!	Strategy General	
Sync up with the medic, who appears to be prioritizing next victim regardless of severity.	Coordination	Explicit Communication: ToM
Red, remember to place a marker block after spotting a victim.	Strategy Sequence	Implicit Communication: Marker
Green, let your teammates know about the critical victim marker you recently placed.	Explicit Communication: Marker	
If you are not sure what to do next, there are currently 9 stabilized victims that need transport.	Information Sharing	Strategy Sequence
Sync up with the medic, who appears to be prioritizing next victim regardless of severity.	Coordination	Explicit Communication: ToM

attributions of the intervention recipient (e.g., "it seems like you…") and second-order attributions of their teammate (e.g., "who appears to be…").

RESULTS

To begin exploring the impact of ASI interventions on team performance outcomes, we focus on the interventions related to the ASI's use of AToM. We specifically look at the code categories labelled a form of communication. This included Explicit Communication: ToM, where the ASI intervention made explicit ToM attributions of a participant's or their teammates' intentions, and Implicit Communication: Marker, where the ASI interventions made suggestions, recommendations or reminders based on inferred participant intentions and actions. Each of the six ASI advisors in this study exhibited different tendencies in their interventions such that all agents delivered interventions of differing frequencies and types, including from the same ASI across different teams. Finally, because we were not interested in human learning, that is, increases in teamwork based upon mission practice, our analyses focus on the 2nd Mission participants completed as part of the study so that the learning effects between the two missions would not confound our results. Further, due to space limitations, analysis of the entirety of the interventions dataset and all categories is not provided.

Count of Interventions

First, we consider overall counts of interventions and, separately, the percent of a given mission's interventions belonging to certain categories to account for those output disparities. Notably, overall count of interventions delivered by an ASI only correlated significantly with one performance outcome metric, the average lag time between discovery of a victim and the rescue of that victim (Pearson's $r = -0.198$, $p = .025$). As can be seen in Table 3, the count and percentage of interventions explicitly referencing agent or team member theory of mind did not correlate positively with mission outcomes.

Table 3. Correlations between explicit communication: theory of mind interventions and mission outcomes.

Variable 1	Variable 2	Pearson's r	p
Intervention Count: Total	Intervention Count: Explicit Theory of Mind	0.489***	<.001
Intervention Count: Total	Intervention Percent: Explicit Theory of Mind	0.045	0.381
Intervention Count: Explicit Theory of Mind	Metric: Mission Score	−0.286	.997
Intervention Count: Explicit Theory of Mind	Metric: Critical Victim Rescues	−0.242	.989
Intervention Percent: Explicit Theory of Mind	Metric: Mission Score	−0.076	.690

Note: all tests one-tailed, for positive correlation. *p<.05, **p<.01, ***p<.001

Explicit Communication: ToM Interventions and Mission Outcomes

Results indicate that the number of interventions making explicit ToM attributions may in fact have negatively impacted teams' outcomes. As can be seen in Table 4, both overall score and critical victim rescues were significantly reduced as the count of theory of mind related interventions increased.

Table 4. Correlations between explicit communication: theory of mind interventions and mission outcomes.

Variable 1	Variable 2	Pearson's r	p
Intervention Count: Explicit Theory of Mind	Metric: Mission Score	−0.286**	.003
Intervention Count: Explicit Theory of Mind	Metric: Critical Victim Rescues	−0.242**	.011

Note: all tests one-tailed, for negative correlation. *p<.05, **p<.01, ***p<.001

DISCUSSION

This paper reports a portion of a larger research program on development of artificial social intelligence. Through analyses of interventions generated by

ASI during team interactions on a complex task, we examined how Artificial Theory of Mind (AToM) influences team outcomes. Our results demonstrate how the explicit statement of ToM attributions made by the agent may be unhelpful to human teammates. Although we were not able to examine why this occurred, it could be that human members of the team developed and maintained their own ToM with regards to the experimental tasks and their teammates. As such, the ASI may have interfered with human social cognitive processes emergent in this complex task. Additionally, it could be due to interventions not being of use for some team. For example, a participant may have found the intervention "Blue, it seems you need some help to rescue a critical victim. Ask your teammates for assistance" (which received a primary code of Explicit Communication: ToM) to be redundant or not useful if the agent was making observations and suggestions the player may have been aware of already. This is particularly the case when seen in contrast to other interventions. For example, the intervention "Sync up with the medic, who appears to be prioritizing next victim regardless of severity" received a primary code of Coordination and a secondary code of Explicit Communication: ToM. This intervention could provide greater value to participants because it is explaining a problem and offering coordinative corrections. Specifically, the ASI both observes teammate actions and predicts intentions while then relaying them to the participant to provide task-relevant knowledge of their teammate and suggesting coordinative actions to support teamwork.

Based upon these analyses, we can offer suggestions for future research in ASI and the implementation of agent architectures capable of AToM. First, research should examine how more sophisticated modelling can provide explicit ToM attributions in interventions that enhance team member situation awareness. Second, research should study if agents better equipped to provide progress and goal monitoring behaviors, improve processes and/or outcomes. Third, research can study if ASI providing coordinative guidance, for example, reminders related to time-sensitive or event-based tasks, improves teamwork. Fourth, research could study how AToM can be used to improve 'explainable AI' (DARPA, 2016; Gunning et al., 2019); that is, can AToM improve an agent's ability to communicate the rationale for decisions, instead of, or in addition, to, making explicit ToM attributions. As the advancement of ASI to support human-agent teaming continues, engaging appropriate and effective interactions/interventions will require the use of AToM, just as expert human-human teams do. And the presentation, structure, and content of these interventions still needs to be further investigated to determine what types of interventions are most effective at influencing different types of teamwork processes and outcomes.

In sum, this study represents one of the first analyses of the direct contact point between human team members and the developing ASI agents. Specifically, the study of AToM is still in its infancy, and the larger research program associated with this study is developing some of the first agent architectures capable of artificial social intelligence. We report on the direct contact between those architectures and their human teammates. As such, this is an important contribution to the human-machine teaming literature. First, it examines ASI actively monitoring humans in human-machine teams. Second, the ASI is designed to directly interact with humans in an advisory role during a dynamic/complex team task. And, third, ASI generates interventions that are tailored to particular performance needs, focusing on teamwork, that are largely based on AToM models.

ACKNOWLEDGMENT

This research was sponsored by the Defense Advanced Research Projects Agency (DARPA) and the Army Research Office and was accomplished under Grant Number W911NF-20-1-0008. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of DARPA, the Army Research Office, or the U.S. Government. The Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

Alonso, V., & De La Puente, P. (2018). System transparency in shared autonomy: a mini review. *Front. Neurorobot.* 12, 83. doi: 10.3389/fnbot.2018.00083.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(2), 241–251.

Bendell, R., Williams, J. Fiore, S. M., & Jentsch, F. (2021). Supporting Social Interactions in Human-AI Teammates from Sparse Data. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (665-669). Los Angeles, CA: SAGE Publications.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, *37*(2), 255–285.

Cooke, N. J., & Shope, S. M. (2017). Designing a synthetic task environment. In *Scaled worlds: Development, validation and applications* (pp. 273–288). Routledge.

Corral, C. C., Tatapudi, K. S., Buchanan, V., Huang, L., & Cooke, N. J. (2021, September). Building a synthetic task environment to support artificial social intelligence research. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 660–664). Sage CA: Los Angeles, CA: SAGE Publications.

Cuzzolin, F., Morelli, A., Cirstea, B., and Sahakian, B. J. (2020). Knowing me, knowing you: theory of mind in AI. *Psychol. Med.* 50, 1057–1061. doi: 10.1017/S0033291720000835.

DARPA (2016). *Explainable Artificial Intelligence (XAI)*. Technical Report Defense Advanced Research Projects Agency.

Demir, M., McNeese, N. J., & Cooke, N. J. (2020). Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies*, *140*, 102436.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of Applied Psychology, 95*(1), 32–53.

Fiore, S. M., Bracken, B., Demir, M., Freeman, J., Lewis, M. (2021). Transdisciplinary Team Research to Develop Theory of Mind in Human-AI Teams. *Proceedings of* *the Human Factors and Ergonomics Society Annual Meeting* (1605–1609). Los Angeles, CA: SAGE Publications.

Fiore, S. M. & Schooler, J. W. (2004). Process mapping and shared cognition: Teamwork and the development of shared problem models. In E. Salas & S. M. Fiore (Editors). *Team Cognition: Understanding the factors that drive process and performance* (pp. 133–152). Washington, DC: American Psychological Association.

Fiore, S. M., Wiltshire, T. J., Lobato, E. J. C., Jentsch, F., Huang, W. H., & Axelrod, B. (2013). Towards Understanding Social Cues and Signals in Human-robot Interaction: Effects of Robot Gaze and Proxemic Behavior. *Frontiers in Psychology: Cognitive Science*. 4:859.

Freeman, J., Huang, L., Wood, M., & Cauffman, S. J. (2023, January). Evaluating artificial social intelligence in an urban search and rescue task environment. In *Computational Theory of Mind for Human-Machine Teams: First International Symposium, ToM for Teams 2021, Virtual Event, November 4–6, 2021, Revised Selected Papers* (pp. 72-84). Cham: Springer Nature Switzerland.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., and Yang, G. Z. (2019). XAI-Explainable artificial intelligence. *Sci. Robot.* 4, 37. doi: 10.1126/scirobotics.aay7120.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, *30*(5), 425–447.

Horst, F., Lapuschkin, S., Samek, W., Müller, K. R., and Schöllhorn, W. I. (2019). Explaining the unique nature of individual gait patterns with deep learning. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-38748-8.

Huang, L., Freeman, J., Cooke, N., Colonna-Romano, J., Wood., M., Buchanan, V., & Caufman, S. J. (2022a). Exercises for Artificial Social Intelligence in Minecraft Search and Rescue for Teams. OSF. https://osf.io/project/jwyvf/files/osfstorage/62549a7c28f940000fa2455e

Huang, L., Freeman, J., Cooke, N., Colonna-Romano, J., Wood, M., Buchanan, V., & Caufman, S. (2022b). *Artificial Social Intelligence for Successful Teams (ASIST) Study 3* (ASU Library Research Data Repository; V3) [Data set, study procedure and materials]. ASU Library Research Data Repository. https://doi.org/10.48349/ASU/QDQ4MH

Joo, H., Simon, T., Cikara, M., and Sheikh, Y. (2019). "Towards social artificial intelligence: Nonverbal social signal prediction in a triadic interaction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 10873–10883.

Li, H., Le, L., Chis, M., Zheng, K., Hughes, D., Lewis, M., & Sycara, K. (2023, January). Sequential theory of mind modeling in team search and rescue tasks. In *Computational Theory of Mind for Human-Machine Teams: First International Symposium, ToM for Teams 2021, Virtual Event, November 4–6, 2021, Revised Selected Papers* (pp. 158–172). Cham: Springer Nature Switzerland.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of applied psychology*, *85*(2), 273.

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors*, *60*(2), 262–273.

Newton, O. B., & Fiore, S. M. (2020, December). What's the Difference? Reconciling Knowledge Structure Concepts to Aid AI Development for Human-Machine

Teaming. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 64, No. 1, pp. 1110–1114). Sage CA: Los Angeles, CA: SAGE Publications.

Oguntola, I., Hughes, D., and Sycara, K. (2021). "Deep interpretable models of theory of mind," in *2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (Vancouver: IEEE), 657–664. doi: 10.1109/RO-MAN50785.2021.9515505.

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Hum. Behav. Emerg. Technol.* 1, 33–36. doi: 10.1002/hbe2.117.

Vinciarelli, A., Pantic, M., Bourlard, H., and Pentland, A. (2008). "Social signals, their function, and automatic analysis: a survey," in *Proceedings of the 10th International Conference on Multimodal Interfaces*, Chania Crete, 61–68. doi: 10.1145/1452392.1452405.

Wang, Y., Gurney, N., Zhou, J., Pynadath, D. V., & Ustun, V. (2023, January). Route Optimization in Service of a Search and Rescue Artificial Social Intelligence Agent. In *Computational Theory of Mind for Human-Machine Teams: First International Symposium, ToM for Teams 2021, Virtual Event, November 4–6, 2021, Revised Selected Papers* (pp. 220–228). Cham: Springer Nature Switzerland.

Williams, J., Fiore, S. M., & Jentsch, F. (2022). Supporting artificial social intelligence with theory of mind. *Frontiers in artificial intelligence, 5*, 750763.

Wiltshire, T. J., Lobato, E. J., Velez, J., Jentsch, F., and Fiore, S. M. (2014). "An interdisciplinary taxonomy of social cues and signals in the service of engineering robotic social intelligence," in *Unmanned Systems Technology XVI* (Baltimore, MA: International Society for Optics and Photonics). doi: 10.1117/12.2049933.