Artificial Intelligence, Social Computing and Wearable Technologies, Vol. 113, 2023, 165–175
https://doi.org/10.54941/ahfe1004189

Interventions by Artificial Socially Intelligent Agents in Collaborative Environments: Impacts on Team Performance and Knowledge Externalization

Rhyse Bendell[1], Jessica Williams[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

The research reported here explores the interactions between AI imbued with Artificial Theory of Mind and teams of human participants completing simulated Urban Search and Rescue missions. The focus of our explorations are the relationships between the advisory interventions delivered by artificial, socially intelligent agents and the mission outcomes of the teams with which they worked. The gamified Urban Search and Rescue task employed for this research consisted of two missions per team during which participants searched for, triaged, and evacuated victims of a building collapse. Each three-person team was assigned an ASI agent who interacted with them during both missions. Of primary interest to this work is the nature of the advisory interventions delivered by the agents while assisting with the rescue missions. Critically, the advice that agents delivered to teams was based entirely on their artificial theory of mind and not rote problem solving. In this paper, we focus on exploring the interventions with attention to the nature of the content and delivery, and a particular interest in the interventions associated with team communication. The results of these analyses suggest that, overall, interventions were generally associated with positive outcomes rather than negative ones. Specifically, interventions advising teams to engage in information sharing and externalizing communication tended to relate positively to outcomes. That finding indicates that even early forms of artificial social intelligence have the potential to serve as teammates as opposed to be utilized as tools, and that artificial teammates can improve team performance. Further, the correlations between communication intervention types and mission performance reflect on how artificial social intelligence can support teams to more effectively engage in teaming activities, such as communication, which can benefit team performance outcomes. These findings are an important step towards investigating the impact of agents actively engaging in teaming behaviors, demonstrating an agent's potential benefit to teamwork by supporting team communication and, additionally, identifying what factors may have negatively impacted performance and should be avoided to improve team effectiveness.

Keywords: Teaming, Human-autonomy teaming, Simulated task environment, Artificial social intelligence

INTRODUCTION

In order for Artificial Intelligences (AI) to take on more teaming and collaborative responsibilities, an AI needs to be able to perceive, extract, and interpret the social information contained in interactions with others (Wiltshire et al., 2014). Thus, an AI will need to be capable of understanding and simulating human mental states, such as their beliefs, intentions, emotions, and motivations. In humans, these abilities are enabled through a core cognitive process known as a Theory of Mind – the implementation of architectures designed to imbue this process in artificial agents is what we refer to as Artificial Theory of Mind (Williams et al., 2022). By using ToM and social intelligence human teams are able to coordinate specialized roles and provide knowledge that anticipates each other's needs. Similarly, an agent will need to be capable of artificial social intelligence (ASI) and AToM in order to better anticipate and predict human needs and respond to humans in ways that are able to take into account various factors in a situation

Here, we look at the interactions between real ASI agents and their human teammates, focusing on the types of interactions given by different ASI. The present research explores these interactions with a focus on the content and delivery of advice stemming from a given artificial agent's theory of mind. The primary question of interest is how artificial teammates can successfully work with teams not simply to conduct taskwork, but also to engage in and support teamwork processes. For example, an agent may have the functionality to calculate an ideal distribution of resources across team members in service of task execution, but to engage as a team member it must also be able to recognize and align the beliefs and goals of its teammates to direct (or, ideally, collaborate on the direction of) sequential and simultaneous interdependent tasking. The explorations reported in this manuscript relate to artificial agents developed under the "Artificial Social Intelligence for Supporting Teams" (ASIST) program (DARPA, 2016), and interactions that those agents had with teams of individuals collaborating to complete a simulated search and rescue task (see Methods). The ASIST agents are an important step towards functional ASI teammates because they were designed with an intentional focus on development of artificial theory of mind, and are the culmination of interdisciplinary efforts to not only develop AI but also to further the methods necessary to test those agents and evaluate team performance in the context of the needs of near-future human-agent teams. The ASIST agents have therefore been tested with real teams in a virtual environment that supports in depth exploration of individual team member performance, team processes, team performance outcomes, and team outcome perceptions in the context of human-agent teaming. The explorations discussed below are one facet of that work which provides insight into the features of agent interactions that may positively improve teamwork, and those that may inhibit successful team performance.

METHODS

Artificial Social Intelligence Supporting Teams: Urban Search and Rescue Task

The analyses discussed below draw upon data collected in an experiment run by Arizona State University as part of the third program experiment of the ASIST; this data has been made publicly available (see: Huang et al., 2022a). The experimental task in this study was a simulated Urban Search and Rescue (USAR) mission using a Minecraft-based testbed (see Huang et al., 2022b). Participants performed the task in teams of three with each individual performing a different role in service of the overarching team goal of rescuing victims from a partially collapsed building. One team member served as a medic who could diagnose injuries, another member as an engineer who could identify the locations of weakened structures that might collapse and endanger the team, and the final member was equipped with a heartbeat sensor for searching the environment for victims. All team members had the ability to transport victims; however, each role was associated with different movement speeds such that the engineer moved quite slowly, the medic at a middling speed, and the searcher almost twice as fast. Each team member's virtual avatar wore a different colored outfit, red green or blue, that was used to distinguish them and served also as a means of communication reference such that, for example, the medic who wore a red shirt could be referred to as 'Red' by teammates. The fourth and final member of each team was one of the Artificial Social Intelligences developed in service of the ASIST program. There were six different agents that were instantiated in the testbed as a disembodied advisor or assistant to the team each made by a different ASIST performer: DOLL, CMU, USC, CRA, and UAZ (for details and documentation interested readers are referred to Huang et al. 2022b). These agents interacted with teams and displayed diverse intervention strategies. These interventions varied in type, frequency, and effectiveness, influencing the teams' performance and perceptions.

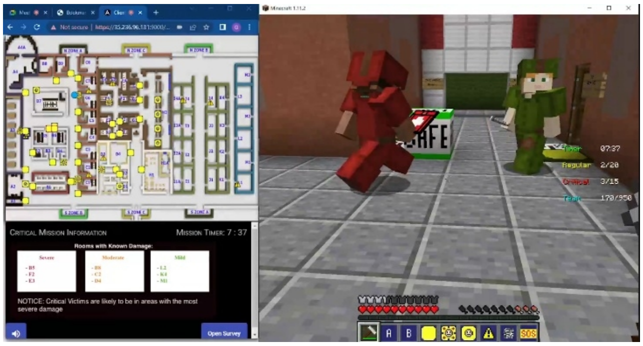

Figure 1: A participant controlling the engineer avatar coordinates with the medic, 'Red,' and the searcher, 'Green,' to stabilize and transport a victim. The left portion of the screen displays the participant's mini-map featuring information gathered and shared by the team.

Teams faced several challenges when completing the search and rescue operations: first they had to coordinate to locate victims for the medic to triage while also managing the risks that were known only to the engineer, and second they had to stabilize those victims and transport them to the extraction point that was appropriate for their injuries. Critically injured victims required additional coordination among the team as two members had to be in close proximity with those victims in order to stabilize them. To support coordination of knowledge and strategies, teams were provided a shared mini-map to which they could externalize information about the mission such as the location of victims, the type of injury sustained by a given victim, the hazardous areas of the building that were likely to collapse, as well as information about which areas they had already searched. Participants were able to save information to the shared mini-map by placing down "markers" in the environment that provided an in-simulation reference as well as automatically placing a symbol on the team's maps. Each mission completed by the team lasted for 17 minutes or until all victims had been successfully extracted to the correct locations. Teams completed two missions.

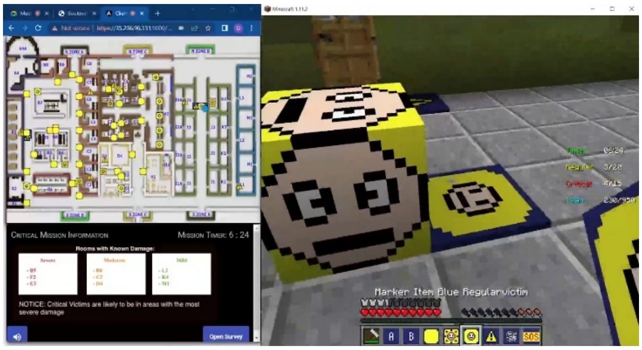

Figure 2: A participant controlling the engineer avatar marks the location of a victim that they have found in the partially collapsed building by placing a 'victim' tile on the ground.

ASI Interventions: Coding Intervention Types

Examining 232 simulated missions, we looked at all of the interactions between the ASI and human teammates, which were implemented in the form of interventions. In total, the ASI advisors delivered 2067 messages to teams to assist them with their completion of the search and rescue. These interventions were times that the ASI gave the human team advice or updates based on the events and actions that happened during their mission. The methods used in the intervention coding process have been described in (Williams et al., 2023) – the reader is referred to this for detail. To summarize these methods briefly, we compiled every unique intervention given and then conducted a reflexive thematic analysis to capture what the goal or intention of the intervention was, so that we could see how humans responded to it – this could range from interactions like "mission status updates" where the ASI updated the team on how many minutes they had remaining in the mission or how many points they had earned through successful rescue, to "motivation" where the ASI cheered them on or told them they did a good job on a task. In this paper, we focused on a subset of the intervention types and the mission outcomes associated with those intervention types in the context of a given agent. Table 1 contains some examples of the types of interventions delivered by ASI advisors and the codes that were applied for the reported explorations.

Table 1. Examples of ASI interventions and associated coding scheme.

Intervention (**agent advice**)	**Code**
Sync up with the medic, who appears to be prioritizing the next victim regardless of severity.	Coordination
Apply more of your medic-specific skills, such as stabilizing victims. Consider grouping more closely with the engineer.	Strategy Role Capability
Keep a special eye out for markers before entering threat rooms! You definitely don't want to waste time getting stuck in a room that was already marked.	Implicit Communication: Marker
'Green' hasn't removed ANY markers yet. obsolete markers can lead to team confusion. I suggest that you ask 'Green' to remove markers of victims before transporting.	Explicit Communication: Strategy
'Blue,' it seems you need some help to rescue a critical victim. Ask your teammates for assistance.	Explicit Communication: ToM

EXPLORATORY FINDINGS

No two agents that interacted with participants in ASIST's study three behaved alike in that the types and frequencies of interventions they delivered varied widely. Partly, that is a natural result of agents partnering with different teams that displayed different needs; however, it also reflects the fundamental differences in the approach taken by each agent in the development of their artificial theory of mind and decision making regarding best strategies for offering support. There was, however, notable consistency within agents regarding the types of interventions they were likely to administer. For example, some agents strongly targeted motivational messages, whereas others delivered no motivational interventions. Similarly, some agents chose to explicitly address theory of mind elements (e.g., sharing beliefs regarding one player with another, referencing their internal belief state when rendering advice, speculating about the goals of a given player in relation to coordination needs) in many of their interventions across teams, but others shifted the profile of their intervention types between the teams with which they worked.

It is important to keep in mind that the analyses described in this section are not aimed at comparing agents to each other, but they are also not agnostic to the agent with which a given team collaborated. Here, we look at the outcomes associated with different intervention types as they manifest within the particular teams that worked with a given agent; details about the architectures and design of each agent can be reviewed at the ASIST dataverse (Huang et al., 2022b), but are largely irrelevant for the present analyses. Any agent in this study could administer any type of intervention as coded using the system described above (see Methods), and their effects on teams may have differed. Accordingly, one agent may have delivered a higher percentage of "Information Sharing" type interventions that was correlated with "Mission Score" (see Table 1) whereas another may have delivered a higher percentage of "Strategy Role Capability" type interventions that was correlated with "Team Resilience Scores" (see Table 2). All intervention variables were calculated in terms of the percent representation of that type of intervention as a function of the total number of interventions delivered over the course of a search and rescue mission. Additionally, due to learning effects that cause teams to perform both better and notably different between their first and second missions we are examining only outcomes related to teams' final missions. Finally it is important for the reader to understand that these are exploratory analyses and a liberal criterion (alpha = 0.05) was applied, which increases the potential for false positives. This approach was selected as we are not intending to test hypotheses, but to explore the interaction space and potential avenues for study.

Table 2. Information sharing, mission status updates, and team outcomes.

Intervention Code (% of total in mission)	Outcome Metric	Pearson's r	p value	Team Agent
Information Sharing	Victim discover-to-extraction lag time	−0.557	.039	CMU
Information Sharing	Mission score	0.535	.049	CRA
Mission status update	Mission score	0.574	.040	USC

Table 3. Strategy related interventions and team outcomes.

Intervention Code (% of total in mission)	Outcome Metric	Pearson's r	p value	Team Agent
Strategy sequence	Victim discover-to-extraction lag time	0.726	.003	DOLL
Strategy role capability	Team resilience scores	0.635	.015	SIFT
Strategy role capability	Mission Score	0.678	.011	SIFT
Strategy role capability	Critically injured victim extractions	0.684	.010	SIFT

Regarding interventions that were focused on delivering mission relevant information and those targeted at team-level strategies: the impacts on team outcomes were largely positive. Particularly variables with straightforward face value such as "Mission score" and "Team resilience scores," the identification of positive correlations with certain intervention types is an important indicator that ASI can have positive effects on team outcomes (see

Table 1, Table 2). We interpret this pattern of findings to suggest that mission information and strategy focused interventions impacted taskwork and task related outcomes somewhat more than teamwork because these interventions dealt mostly with rote, game mechanic-based information and impacted task related outcomes.

There is a pair of somewhat conflicting findings represented in these tables related to victim rescue lag time. First, we found a negative correlation between "Information sharing" and the lag between discovery of victims and successful rescue of those victims meaning that teams receiving more "Information sharing" interventions demonstrated less lag time. Second, a positive correlation between "Strategy sequence" and lag time indicates that teams receiving more "Strategy sequence" interventions took longer between locating and evacuating victims. A full review of the interdependent tasks used to test the human-agent teams (out of scope here, but reviewable in Huang et al., 2022b) shows that an increase in this lag time is not necessarily a negative outcome, but may in fact indicate greater strategy coordination among teams. Particularly, because this metric measures lag from discovery to extraction it likely indicates that teams engaged in a "extraction burst" strategy, meaning that they first located, stabilized, and then placed victims in a central location before transporting them en masse towards the end of the mission. Accordingly, the negative correlation between "Information sharing" and lag time may be viewed at a high level as a marginally bad outcome; however, outside of the context of individual team performance it is irrelevant to assume one way or the other. What is especially important to note with these findings is that there were differences in lag time outcomes that were associated with the types of interventions that ASI were delivering, and that one type of intervention does not necessarily follow the same pattern of relations as another.

Table 4. Motivation interventions and team outcomes.

Intervention Code (% of total in mission)	Outcome Metric	Pearson's r	p value	Team Agent
Motivation	Mission score	-0.561	.037	CRA
Motivation	Victim discover-to-extraction lag time	-0.675	.008	DOLL
Motivation	Critically injured victim extractions	-0.607	.028	USC

"Motivation" focused interventions seemed to have a negative relationship with team outcomes for half of the ASIST agents (see Table 3). Mission score and critically injured victim extractions both reduced as agents delivered more "Motivation" related interventions. As discussed above, there may be some argument to be made that reduction of lag time is not necessarily a negative result; however, in the context of the other findings we interpret this to indicate that teams were not engaging in an overarching coordinated strategy but were primarily conducting individual taskwork as the work became available. These findings are interesting because there is some evidence in human teams that motivational factors can have a positive impact on teaming; however, that does not seem to translate readily into messages sent by artificial agents. It is possible that this finding is partly a result of the poor construction of many of the motivation messages delivered by agents in this study. “Keep going! Do your best teamwork” was the most common intervention in the “Motivation” category, which unfortunately does not relate to a given team’s particular performance or status. These sorts of inadequately tuned, impersonal interventions may have simply distracted teams or in the worst case may have annoyed them and caused them to exert less effort. There were, however, more tailored “Motivation” focused messages such as “Your team is doing great. Compared with other teams you are doing better than average, so keep doing what you are doing.” – a notable failing of our current approach to exploring this data is an inability to separate out the effects of such targeted interventions from the more generic messages.

Table 5. Intervention types related to team knowledge externalization behaviors.

Intervention Code (% of total in mission)	Outcome Metric	Pearson’s r	p value	Agent
Explicit Communication: Theory of Mind	Markers placed	0.595	.025	CRA
Coordination	Markers placed	0.660	.010	DOLL
Strategy: general	Markers placed	-0.555	.039	SIFT
Implicit communication: markers	Markers placed	-0.599	.025	USC
Motivation	Markers placed	-0.658	.011	CRA
Explicit Communication: Theory of Mind	Markers removed	0.705	.005	SIFT
Strategy: sequence	Markers removed	0.626	.017	CRA
Strategy: role capability	Markers removed	-0.579	.030	CRA
Implicit communication: markers	Markers removed	-0.589	.027	USC
Motivation	Markers removed	-0.697	.006	CRA

Table 4, above, is organized slightly differently than the previous tables in that we have clustered outcomes related to knowledge externalization behaviors as opposed to clustering based on the type of intervention. Knowledge externalization behaviors, which were foundational for information sharing and coordination, differed widely amongst teams and showed remarkably mixed relationships across ASIs. Placement of marker blocks was the primary knowledge sharing mechanism available to team members and was important for indicating the locations of mission objectives for coordination. Separately, removal of marker blocks was critical for updating and maintaining team level knowledge so that individual members could increase the efficiency of their actions and the team could coordinate based on relevant and current information. Not all teams employed the marker blocks as intended with some teams essentially neglecting to use the markers and others using them to mark their paths through the environment rather than to mark knowledge gathered from the space; however, overall most teams used the blocks as intended.

Intervention types had a mix of relations to the usage of marker blocks. Regarding placement of blocks, interventions related to “Explicit communication: theory of mind” and “Coordination” were associated with an increase in the placement of blocks. Interestingly, "Strategy: general," "Implicit communication: marker blocks" and "Motivation" interventions led to a reduction in the usage of marker blocks. Although the negative impact of motivational interventions is not surprising in the context of the findings in Table 3, it is notable that "Implicit communication: marker blocks" led to less usage considering that these interventions were directly targeted at improving marker block usage.

Removal of marker blocks followed a similar pattern to placement in that "Explicit communication: theory of mind" and "Strategy: sequence" correlated with more updating and maintenance of marker blocks whereas "Implicit communication: markers," "Motivation," and "Strategy: role capability" were associated with less. We interpret this to suggest that teams which responded to certain types of interventions modified their behaviors with respect to marker blocks for both placement and removal.

CONCLUSION

Artificial teammates will be an integral part of future teams across many domains. The success of those teams will be dependent on the abilities of all team members, human and artificial, to successfully execute taskwork as well as to engage in the collaborative process of teamwork. Here, we have examined the interactions between early forms of such human-agent teams, which included artificial agents imbued with theory of mind and social abilities to observe, understand, and intervene with advice for their human counterparts. Our analysis placed extra emphasis on the nature of the particular interaction events so as to support an exploration of the features of those interactions that positively and negatively related to team processes and team outcomes. Overall, we found that the interventions delivered by the artificial social intelligences (ASI) that interacted with virtually instantiated search and rescue teams were associated with positive outcomes. Mission scores, team resilience outcomes, interdependent task coordination, and knowledge externalization/maintenance outcomes were all found to have primarily positive relationships with different types of ASI interventions. Those relationships did, however, differ across the types of interventions as distinguished by the application of a coding scheme targeted at identifying the content and artificial theory of mind (AToM) basis of interventions. Intervention types such as those focused on "Information sharing" and "Strategy sequence" or "Coordination" content showed promisingly positive relationships with mission outcomes whereas "Motivation" interventions were found to have largely negative relationships with outcomes such as team's mission scores and knowledge sharing behaviors. Further, types of interventions did not necessarily uniformly impact outcomes such that "Strategy: role capability" interventions, for example, positively related to mission score and team resilience score outcomes but negatively related to appropriate use of knowledge sharing capabilities.

These findings provide important insights into the outcomes one of the first studies in which ASI agents with social capacities interacted in real-time to observe human team members, understand those member's beliefs and goals, reason meaningfully about the team processes needed to coordinate and achieve those goals, and interact with their human teammates to support the accomplishment of team objectives. Our insights demonstrate that socially intelligent artificial teammates can successfully work with human teammates and intervene in ways that are associated with positive team outcomes across multiple metrics. Further, we have highlighted that there are important differences related to the type of intervention content delivered by those agents, and that those differences may manifest differently for ASI-to-human interactions than they do for human-to-human interactions. Motivational interventions in particular seemed to have only negative impacts on team members and team processes, which may have been a harmful result exacerbated by the fact that those messages were being delivered by a disembodied artificial agent as opposed to a human.

Future examinations of human-agent teams that attend to the social aspects of their interactions should be sure to consider the following two issues: it is critical to distinguish interactions between teammates at the event level because some may be similar and therefore related but many must be evaluated separately (consider, for example, the differences between strategy related and motivation related interactions in the above analyses), and second it is important to consider the impact of having an artificial agent in the loop. To the first issue, one of the limitations of the explorations reported here is that we cannot currently examine outcomes associated with individual interventions (these efforts are ongoing), but we can already recognize at the aggregate level that distinguishing between the features of those events is necessary for further understanding. To the second issue, it is currently unknown what features of artificial teammates will impact the social space that has till now been inhabited only by human actors. The future of human-agent teams research will benefit from consideration of what types of interactions artificial social agents can use to impact teams positively and which to avoid (or use with caution) in the service of team goals and successful team processes.

ACKNOWLEDGEMENTS

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

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

Huang, L., Freeman, J., Cooke, N., Colonna-Romano, J., Wood, M., Buchanan, V., & Caufman, S. (2022a). 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

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

Williams, J., Fiore, S. M., & Jentsch, F. (2022). Supporting Artificial Social Intelligence with Theory of Mind. *Frontiers in Artificial Intelligence, 5,* 750763.

Williams, J., Bendell, R., Fiore, S. M., & Jentsch, F. (2023). The Role of Artificial Theory of Mind in Supporting Human-Agent Teaming Interactions. In Proceedings of the AHFE 2023 International Conferences on Human Factors and Simulation held on July 20–24, 2023, at San Francisco Marriott Marquis, San Francisco, California, USA. Springer International Publishing.

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.