INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2022.2085191

Investigating the Effects of Perceived Teammate Artificiality on Human Performance and Cognition

Beau G. Schelble[a], Christopher Flathmann[a], Nathan J. McNeese[a], Thomas O'Neill[b], Richard Pak[c], and Moses Namara[a]

[a]Division of Human-Centered Computing, Clemson University School of Computing, Clemson, SC, USA; [b]Department of Psychology, University of Calgary, Calgary, Canada; [c]Department of Psychology, Clemson University, Clemson, SC, USA

ABSTRACT

Teammates powered by artificial intelligence (AI) are becoming more prevalent and capable in their abilities as a teammate. While these teammates have great potential in improving team performance, empirical work that explores the impacts of these teammates on the humans they work with is still in its infancy. Thus, this study explores how the inclusion of AI teammates impacts both the performative abilities of human-AI teams in addition to the perceptions those humans form. The current study found that participants perceiving their third teammate as artificial performed worse than those perceiving them as human. Furthermore, these performance differences were significantly moderated by the task's difficulty, with participants in the AI teammate condition significantly outperforming participants perceiving a human teammate in the highest difficulty task, which diverges from previous human-AI teaming literature. Alternatively, no significant effect of perceived teammate artificiality was found on shared mental model similarity. However, it did significantly affect participants' levels of perceived team cognition. Individual performance on medium difficulty maps also mediated the effect of perceived teammate artificiality on perceived team cognition. These results further build on the current understanding of how AI teammates impact perceptions of individual human teammates and how those perceptions subsequently impact their objective performance, which is critical in building more effective AI teammates to incorporate alongside humans.

1. Introduction

Human-AI teaming is rapidly becoming a reality for modern workforces and governments to deploy in applied settings to enhance the outputs of their teams (Allam & Dhunny, 2019; Schelble et al., 2020). Recent advances in AI technology like open source reinforcement learning libraries (Schaarschmidt et al., 2017) have begun to give AI-powered artificial agents the ability to operate as viable teammates for work with humans (McNeese et al., 2018). To be classified as a true human-AI team, the AI must operate as an independent entity, be given a significant degree of agency within the team, and work interdependently alongside their human teammates towards a common goal (O'Neill et al., 2020). The Industry 4.0 movement in manufacturing is an excellent example of an applied sector targeting greater AI involvement within the workforce to enhance adaptability, efficiency, and safety (Mabkhot et al., 2018; Schelble et al., 2020). As such, understanding the constructs that apply to human-AI teams and the characteristics that drive effective and ineffective teaming has become paramount to human-AI teaming research (O'Neill et al., 2020).

Several studies have discovered several constructs and factors that significantly influence human-AI teaming effectiveness. This compendium of recent studies has displayed that human-AI teams can outperform human-only teams (McNeese et al., 2018) and has linked human-AI team performance to factors like trust (McNeese et al., 2021), transparency (Mercado et al., 2016), and reliability (Chen & Barnes, 2012). From this research, an interesting trend has been identified that is tying team member perceptions to performance (i.e., a distinguishable period over which performance accrues and feedback is available (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers et al., 2000, p. 273)). This burgeoning focus on perception and performance is gaining attention as a major developmental challenge for AI (Kaminka, 2013; Van Riemsdijk et al., 2015) as the advancements in AI technology are beginning to place them in social situations (like teaming) more than ever before, especially as virtual agents (Chaves & Gerosa, 2021; Kepuska & Bohouta, 2018).

This research has highlighted that successfully joining humans and AI together in teams is not holistically dependent on AIs technical ability but also on a myriad of human factors that impact human-AI collaboration in human-AI teams (O'Neill et al., 2020; Shneiderman, 2020). For instance, simply perceiving that a teammate is artificial has been linked to several adverse effects on team processes and

CONTACT Beau G. Schelble bschelb@g.clemson.edu Division of Human-Centered Computing, Clemson University School of Computing, Clemson, SC, USA

outcomes (Musick et al., 2021; Walliser et al., 2017), including performance (Demir et al., 2018). Additionally, task difficulty is known to moderate the influence of teammate perception on performance, as overall team homogeneity has a known relationship with task difficulty in human-human teams (Bowers et al., 2000). At the same time, previous human-AI and human-automation teaming literature has found that human-AI teams suffer in higher task difficulties compared to human-human teams (Fan & Yen, 2010; Wright & Kaber, 2005). These studies highlight how AI can be a heavily perceived concept that can even be subject to a user's previous experiences and the specific context of the task (Chen et al., 2011; Hafizoglu & Sen, 2018a, 2018b). However, there is very little understanding of the mechanism and nature of how perceptions influence human team members' performance, which makes the need for a robust empirical study to investigate these concepts clear.

Importantly, this work also provides further evidence of the impacts of perceiving a teammate as an AI; however, this investigation is scoped beyond simple task performance and looks at other team-specific perceptions that can impact team effectiveness. Specifically, the broader construct of team cognition, often operationalized through shared mental models, maintains a well-established and veritable relationship to team performance (Mathieu et al., 2000). Teams with a high degree of similarity in their mental models are known to execute at a higher level of performance than teams with lower levels of similarity, acting more efficiently to produce high-quality team outcomes. As such, the quality of shared mental models can be considered an aspect of team performance worthwhile to study the influence that team member perceptions have on them, especially in human-AI teams. This assertion is bolstered by the fact that team member perceptions exist under the umbrella of team cognition alongside shared mental models (Cooke et al., 2000; Klimoski & Mohammed, 1994), and the lack of extensive research on shared mental models in human-AI teaming in the past (Hanna & Richards, 2018).

Closing the above-identified research gap will become more critical as human-AI teams are continuously placed in applied settings that run a broad spectrum of environmental factors that change the number of AI teammates and the difficulty of their tasks (Schelble et al., 2020). AI is a complex and subjective concept, and individuals can vary significantly in their knowledge and past experiences with it. Understanding how these perceptions influence human-AI teams is necessary to developing future human-AI teams that work best for all potential users. Human-computer interaction (HCI) is ideally suited to address this research gap given the unique position of user perceptions, which can be influenced by both the technical and humanistic aspects of human-AI teaming. Given the novelty and importance of perceived teammate artificiality on human task performance and task difficulty, this research addresses the following research questions:

1. ***RQ1:*** *How does the perception of working with an AI teammate, compared to working with a human teammate, influence human-task performance based on the difficulty of the task?*
2. ***RQ2:*** *How does the perception of working with an AI teammate, compared to working with a human teammate, influence overall team cognition?*

The current study addresses these research questions using an experimental study with two conditions of perceived teammate artificiality through three stages of task difficulty. Each condition of perceived teammate artificiality completed fifteen rounds of the human-AI task simulation known as "IIHAT" (Implicit Interaction for Human-Autonomy Teams) (Musick et al., 2021). The study's quantitative findings display the effects of perceived teammate artificiality and its interaction with task difficulty on human task performance while highlighting the effect of perceived teammate artificiality on the development of shared mental models and perceived team cognition.

This research contributes to HCI research on human-AI teams in three significant ways: (1) advancing the forward-facing research on human-AI teams within the context of HCI by investigating the relationship of perceived teammate artificiality on human task performance and shared mental model development; (2) building upon the existing HCI research on the effects of perceptions on team outcomes like task performance within environments involving interaction with an AI teammate; (3) providing valuable conclusions on how the effect of human teammates perception of an AI teammate on human task performance is moderated by task difficulty changes.

2. Related work

2.1. Human-AI teaming and its challenges

Human-AI teaming is rapidly becoming a major research domain (O'Neill et al., 2020) as researchers work to get ahead of the massive rollout of applied human-AI teams into the workforce (Allam & Dhunny, 2019; Schelble et al., 2020; Wilson & Daugherty, 2018). Such teams, are defined by an AI driven artificial entity working with one or more human teammates towards a common goal from a unique role within the team that is interdependent upon the other teammates, all while maintaining a significant degree of self-agency (O'Neill et al., 2020). These AI should be seen as full-fledged teammates and not merely as simple "tools" at the team's disposal. Human-AI teams operating according to these standards have not only been shown to be possible in past research but have also been shown to be capable outperforming traditional human-human teams (McNeese et al., 2018).

The potential for human-AI teams is extensive, with applications ranging from city management to applied manufacturing and cyber security (Allam & Dhunny, 2019; Mabkhot et al., 2018). Many of the strengths offered by human-AI teams come from humans' and agents' ability to compensate for each other's weaknesses. For example, agents' are superior at quickly analyzing large data sets and providing detailed information on said data. However, humans are better at taking that information and making sense of it, identifying potential biases, and using intuition and experience to reach higher order conclusions (Dorton & Hall, 2021). The potential for taking advantage of these strengths has made human-AI teaming and collaboration a significant emerging research requirement in HCI (Stephanidis et al., 2019). However, the differences between human and AI teammates are still stark and can cause problems, with communication ability being the most obvious example. AI agents are significantly lacking in their ability to engage in natural language communication with humans (Young et al., 2018), which has hindered human-AI teams' ability to outperform human-human teams because of lackluster information sharing (O'Neill et al., 2020). Human teammates are also aware of these differences and weaknesses, actively making fewer attempts to engage in communication of any kind and indicating lower levels of team cognition (Musick et al., 2021).

Unfortunately, there has been a plethora of evidence to support the assertion that humans hold a negative perception of AI teammates and even AI agents in general. For example, merely believing that the pilot of an unmanned aerial vehicle was autonomous made it more difficult for those teams to plan with one another (Demir et al., 2018), and trusted AI significantly less than a human doctor when receiving medical advice (Yokoi et al., 2021). Humans have also been found to adopt a neo-feudalistic view of AI teammates (i.e., unequal rights) when playing games with them (Wehbe et al., 2017). Such negative perceptions are not entirely surprising as it has been found that humans expect the same kind of teaming behaviors that they would exhibit, like shared understanding and comprehensive communication (Tokadl i et al., 2021; Zhang et al., 2021). AI teammates in their current state are unable to reciprocate many of these advanced team functions, which presents a frustrating challenge for human teammates attempting to work with one or several AI teammates. Human teammates' frustrations can then affect team interaction patterns, unfavorably altering team processes and outcomes (Fiore et al., 2001; Resick et al., 2010), which makes it difficult for human-AI teams to reach their full potential. Characterizing the influence of AI teammate perceptions on teaming outcomes like performance and processes like team cognition is necessary to develop more effective human-AI teams. Specifically, using that information to inform the development of better human-AI team training and AI design.

2.1.1. Team cognition in Human-AI teams

Team cognition is another construct of teaming related to team outcomes, which the current paper analyzes as an aspect of performance. Team cognition encompasses the is an emergent team construct that is developed and shaped by processes that include learning, planning, reasoning, decision making, problem-solving, remembering, implicit coordination, and assessing situations (Cooke et al., 2013; McNeese et al., 2021). The construct of team cognition is traditionally studied and operationalized through two approaches: (1) knowledge (e.g., shared mental models; (Converse et al., 1993)), and (2) ecological interaction (e.g., interactive team cognition; (Cooke et al., 2013)). Most research is on the former, wherein shared mental models involve" an organized understanding or mental representation of knowledge that is shared by team members" (Mathieu et al., 2005) (p. 38). Naturally, to be effective, team members must develop both task and team mental models, which involve understanding both the system and team in which they operate, respectively (Converse et al., 1993). The current study includes shared mental models as an aspect of performance as they are known to promote efficient team coordination and action sequences in human-human teams (LePine et al., 2008; Stout et al., 2017), which is the predominant application of agent teammates in previous teamwork studies (Chen et al., 2017; McNeese et al., 2021).

As an emergent team process, there is empirical evidence demonstrating that changes in team composition can have significant influences on aspects of team cognition (Gorman & Cooke, 2011). Specifically, mixed composition teams had longer communication periods, which is indicative of more effective team cognition in the interactive team cognition model (Gorman & Cooke, 2011). Considering the nature of AI's disadvantages as a teammate (e.g., unable to understand natural language processing and develop shared mental models with human teammates by themselves), team cognition is likely influenced in some respect with the addition of AI teammates. The limited existing research on team cognition in human-AI teams has found that team cognition in human-AI teams benefits greatly from multiple communication modalities (i.e., verbal and non-verbal) (Hanna & Richards, 2018) and when AI is made aware of their human teammates' current cognitive load (Fan & Yen, 2010). Additionally, human-AI teams have tended to exhibit more inconsistency in their shared team mental model similarity than human-human teams (Schelble et al., 2022). These studies emphasize the altered nature of team cognition in human-AI teams, such that team cognition and its development over time may not be exactly the same as in human-human teams. Given the significant relationship that team cognition is known to have with several team processes and outcomes, including performance (Mathieu et al., 2000; Mohammed et al., 2010), understanding team cognition in human-AI teams will contribute to more effective human-AI teams with efficient coordination and communication.

2.2. The influence of perception within human-AI teams

Perceptions in human-AI teams are also essential to consider in the current state of human-AI teaming research because of their significant effect on typical teaming functions like backup behaviors, coordination, and anticipation (Demir et al., 2018; Hanna & Richards, 2018; Musick et al., 2021). Perceptions can change the way human team members see themselves, their teammates, and the team as a whole, subsequently affecting their willingness to engage in cooperative behavior (Walliser et al., 2017). Accordingly, teams with individuals carrying negative perceptions of the team as a whole, its goal, or other teammates will lead to problematic execution of action stage tasks (Demir et al., 2018; Schmidt et al., 2014; Stout et al., 1999, 2017). The action stage is a key area of teaming and involves processes like monitoring, coordination, and backup behaviors, which are all essential to effective teamwork and taskwork (Marks et al., 2001). This inability to efficiently execute during action stage tasks directly contributes to the reduction in performance for some human-AI teams, as seen in Demir and colleagues 2018 work (Demir et al., 2018).

Introducing AI teammates affects human teammates through individual-level factors, impacting their abilities and work styles (Resick et al., 2010). This phenomenon is not uncommon as traditional human-human teams are also influenced by changes to team composition. For example, team composition can influence situational awareness (Gorman et al., 2006), team performance (Spotts & Chelte, 2005), and even team cognition (Gorman & Cooke, 2011). However, the effects of team composition on these team processes and outcomes can be dependent upon the difficulty of the task the team must overcome to achieve their shared goal (Bowers et al., 2000). Specifically, teams that are homogeneous in composition (i.e., gender, attitude, ability, personality) tend to perform better in lower difficulty tasks, but heterogeneous teams typically attain higher performance in higher difficulty tasks (Bowers et al., 2000). It is unknown how task difficulty interacts with perceived teammate artificiality in human-AI teams and if evidence from traditional human-human teams in this area is even applicable. This question is important to answer as human-AI teams are predicted to operate across a vast spectrum of situational contexts of varying difficulty (Allam & Dhunny, 2019; Schelble et al., 2020; Wilson & Daugherty, 2018).

3. Methods

The current research study reports on an experiment in which a triad team of all-humans and a triad team of two humans and a single agent (implemented using the Wizard of Oz (WoZ) approach) coordinated together while engaging in a proprietary task simulation known as "IIHAT" (Implicit Interaction for Human-Autonomy Teams), developed explicitly for human-agent teaming research. The WoZ methodology simulates an AI agent using a human confederate, unbeknownst to the human participants involved in the experiment (Kelley, 1983, 2018). This design effectively isolates the effects of perceptions of working with an AI agent, allowing for a more compelling examination of how perceptions may affect team-related variables. Additionally, teams of three were chosen as they have a few advantages over dyad teams when studying complex team constructs like team cognition due to the increased level of complex group collaboration (Amon et al., Moreland, 2010; Williams, 2010). Triad teams are also widespread in other human-AI teaming research as they represent the lowest number of participants necessary to achieve those complex group interactions (Demir et al., 2018; McNeese et al., 2018; Musick et al., 2021).

Table 1. Experimental conditions.

Condition number	Team composition	Teams
Condition 1 (HHH)	Human-Human-Human	15
Condition 2 (HHA)	Human-Human-Agent	15

Participants were not allowed to communicate verbally or textually during this experiment. The choice to disallow communication was made to isolate the effects of implicit coordination (Entin & Serfaty, 1999; Hanna & Richards, 2014, 2015; Shively et al., 2017) in human-human teams vs. human-agent teams and to ensure participants believed the agent was a true AI. Additionally, a trend towards focusing on implicit communication has received attention in applied human-machine research (Aubert et al., 2018), where implicit communication is defined as any action utilized to convey an agent's planned intentions (does not verbalize or use written language). The study implemented two experimental conditions for a 1×2 design with perceived teammate artificiality being the independent variable (see Table 1). Accordingly, this experimental design allows the effects of perceived teammate artificiality on human team processes and outcomes to be effectively studied.

3.1. Participants

Seventy-five undergraduate participants were recruited from the subject pool of a large university in the United States, resulting in 30 total teams completing the experiment. Fifteen teams participated in each condition. Each team completed the hour and a half session and received course credit as an incentive for their participation.

3.2. IIHAT Simulation task

Participants completed a simulated team task called "IIHAT" (Implicit Interaction for Human-Autonomy Teams), which has been used in past research (Musick et al., 2021). This task was developed explicitly for human-agent teaming research and did not allow communication between players to isolate the effect of implicit communication on team cognition and human-agent teams (Entin & Serfaty, 1999; Hanna & Richards, 2014, 2015; Shively et al., 2017). This control on communication also ensured that the participants did not suspect the AI teammate as a human given the difficulty that AI has with natural language processing (Young et al., 2018).

The goal of the IIHAT simulation is to collect six objectives from a fictitious island to escape in as few moves as possible. The simulation was developed for three players with a player beginning in the top left, bottom left, and bottom right of each map as seen in Figure 1. The simulation introduced interdependence between the players by giving each player a unique ability to access a restricted area of the map where objectives were located. The simulation displayed the map in real-time with the objectives' locations, all three players' locations, restricted map areas, and an information panel on the far right conveying important information like the team's current number of collective moves. The number

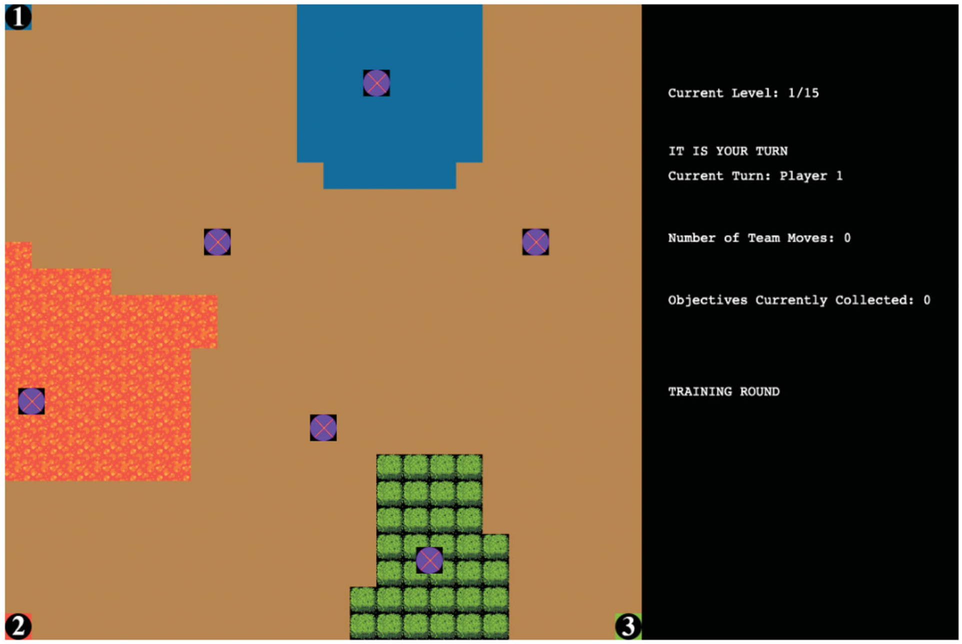

Figure 1. The interface of the IIHAT simulation.

of moves increases as each team member makes their move during their respective turn; however, a team member may skip their turn by pressing T. Turns were taken in numerical order going from Player 1, Player 2, Player 3, and then reverted to Player 1 to repeat the cycle.

Each map developed for the IIHAT simulation contained six objectives for teams to collect before continuing. Three objectives were located in a restricted area and had to be collected by the player with the corresponding ability, while the other three were located in a neutral brown area accessible to all players. The objectives located in the neutral area were placed equidistant between each pair of players in order to encourage interaction and implicit coordination between team members (Entin & Serfaty, 1999; Hanna & Richards, 2014, 2015). After pilot testing the simulation with participants, three-second pauses to gameplay after every third turn and a five-second pause at the start of each map were implemented to slow the pace of play, allowing players to analyze past action and plan a future strategy for the new game state.

Finally, the IIHAT simulation in the current experiment consisted of fifteen unique maps. The first five maps were "easy" difficulty, the second five maps were "medium" difficulty, and the final five maps were "hard" difficulty. Higher difficulty maps necessitated more advanced teammate coordination through additional interaction and interdependence. As such, the IIHAT simulation was developed to ensure high levels of teammate interdependence by presenting a shared goal to each team member to accomplish using the implicit non-verbal communication necessary to carry out a coordinated strategy and develop effective shared mental models (Saavedra et al., 1993; Ven et al., 1976).

3.3. AI teammate

The AI agent was implemented using the Wizard of Oz (WoZ) approach, which simulates the agent's actions while completing the simulation with the participants. The WoZ methodology uses trained human confederates to portray a feature of technology (AI/communication) that participants will see and believe is genuinely the technology's behavior/communication (Kelley, 1983, 2018). Two trained experimenters took the agent's role following a pre-defined protocol to maintain a consistent performance between teams while retaining the ability to react appropriately to the actions and decisions of the participant teammates. This protocol allowed the WoZ agent to maintain extremely high levels of consistency between teams with a standard deviation of 1.2 moves across all fifteen maps and teams. A WoZ agent was not used in the HHH condition to improve the ecological validity of the all-human condition (by allowing for a real human teammate).

The agent teammate was described to participants as an AI trained to complete the task as efficiently as possible hosted on an online database. No other information was given to the participants about the agent's ability to complete the task. This control on information was done to ensure the opinions they formed throughout the simulation were based on their own experiences and not affected by the experimental protocol.

3.4. Procedure

Participants arrived at the study session they signed up for and provided informed consent prior to participating in the experiment. This study was reviewed and approved by the

Clemson University Institutional Review Board under approval number 2019-280. Participants were randomly assigned to a condition and completed the pre-task survey. The pre-task survey collected demographic information and participants' general acceptance of AI. Once the pre-task survey had been completed, the participants received a handout describing the IIHAT simulation task, scenario, and instructions. Participants were also told that they could not communicate textually or verbally throughout the IIHAT simulation task but were primed that implicit communication through action and inaction could convey intention. Before beginning, the participants in the HHH condition were told they would be working together to escape from a deserted island by collecting objectives as efficiently (in as few moves) as possible. Objectives were presented as highly visible and simple circles with a large X painted over them as seen in Figure 1 collected automatically by moving the character into the same tile as the objective. HHA teams were given the same instructions except that they were repeatedly told that they would be working with an AI agent teammate as their third teammate (Player 3). The IIHAT simulation task took roughly twenty-five minutes to complete in its entirety (prior studies have shown this amount of time to be adequate to develop team cognition (McNeese et al., 2021; Musick et al., 2021). Once the IIHAT simulation task was completed, the participants completed a post-task survey that collected information on their task and team mental models, perceived team cognition, and perceived team performance. Finally, a researcher was on hand to ensure the instructions were followed throughout the experiment.

3.5. Measures

3.5.1. Individual human task performance

Human task performance was measured by taking the average of all human teammates' moves and did not include the moves made by the WoZ confederate in the HHA teams, allowing the study to strictly examine only the human teammates response to working with a perceived AI teammate. Human task performance was calculated in this manner for all fifteen maps, and overall human task performance was calculated as the average for all maps, and each level of difficulty was calculated by taking the average for the respective set of five maps. A lower number of moves represents more efficient performance.

3.5.2. Task and team mental models

Mental models were elicited from participants using paired sentence comparison (Bradley & Terry, 1952; Mathieu et al., 2000). This method was chosen due to its reliability and recommendation for use based on its ability to elicit mental models (Mohammed et al., 2010). The method follows the procedure and examples outlined in the seminal paper by Mathieu and colleagues and expounded upon by others (DeChurch & Mesmer-Magnus, 2010; Langan-Fox et al., 2000; Mathieu et al., 2000). Each participant was tasked with comparing the relationships between various task and team-related attributes as positively related (more of one requires more of the other), not related, or negatively related (more of one requires less of the other). For the task mental model, the attributes were generated by conducting a detailed task analysis with two subject matter experts (simulation designers): *(a) amount of information, (b) quality of information, (c) role/responsibility, (d) interaction patterns, (e) communication channels, (f) role interdependencies, (g) information flow, (h) teammates' knowledge, (i) teammates' skill, (j) teammates' attitudes, (k) teammates' preferences, (l) teammates' tendencies*. The attributes team mental model attributes came from past literature (Lee & Johnson, 2008; Mathieu et al., 2000): *(a) identify location of objectives, (b) identify the objective you are responsible for, (c) predict which paths your teammates will take, (d) plan a path to your own objective, (e) determine if extra objectives are on your path, (f) adapt to your teammates movements.*

3.5.3. Mental model similarity

Mental model similarity within a team was measured using the Pathfinder network-scaling algorithm (Schvaneveldt, 1990), a common practice in shared mental model research (Cooke et al., 2003; Mohammed et al., 2000, 2010; Stout et al., 1999). This program uses the matrices produced by the participants' pair-wise comparisons to create graphical representations of mental model networks. A node represents each concept, and the assessed relationship represents each link between the nodes. The resulting comparisons were entered into Pathfinder, which then provided a network similarity score between each dyad within a team, ranging from 0 (indicating no similarity) to 1 (indicating perfect similarity). Teams of three humans had each dyad assessed and then averaged, a common practice in Pathfinder/UCINET shared mental model research (Lim & Klein, 2006; Santos et al., 2015).

3.6. Perceived team cognition

Perceptions of team cognition. Measuring the perceived level of team cognition within teams and the perceived level of team cognition with humans and AI agents was completed using the Teamwork Schema Questionnaire (Pape, 1998; Rentsch et al., 1998). The survey-based measure asked participants to rate a series of statements based on how important each statement was to their idea of teamwork. They then completed the measure again but were asked to assess what they thought each statement meant to their teammate's idea of teamwork (participants assessed human and AI agent teammates separately). This series of questionnaires allowed a measure of congruence to be assessed by taking the absolute difference (Cronbach & Gleser, 1953) between participants' own opinions of teamwork and their assessment of their teammate's opinions on teamwork. Scores were averaged by the number of comparisons made on the team, placing all scores for this measure from 0 to 84, with lower values indicating higher levels of perceived team cognition.

Table 2. Mean and standard deviations for dependent variables.

	HHH		HHA	
Measure	Mean (*N*)	*SD*	Mean (*N*)	*SD*
Total Human Task Performance	357.73 (15)	4.70	362.93 (15)	5.05
Easy Map Human Task Performance	106.27 (15)	2.30	110.90 (15)	1.40
Medium Map Human Task Performance	118.84 (15)	2.75	124.43 (15)	3.70
Hard Map Human Task Performance	132.98 (15)	3.30	127.80 (15)	2.08
Perceived Team Cognition	7.02 (15)	3.07	10.13 (15)	4.77
Team Mental Model Similarity	0.31 (15)	0.09	0.26 (15)	0.09
Task Mental Model Similarity	0.50 (15)	0.13	0.52 (15)	0.13

3.7. General acceptance of AI

Because the current study centered around participants' perceptions of artificial teammates, their general acceptance of AI was taken to act as a covariate when necessary. General acceptance of AI was measured using a modified version of the UTAUT (Pontiggia & Virili, 2010), which replaced the term "technology" with "AI."

4. Results

The following quantitative findings are split into two sections, the first focusing on objective human task performance and the second focusing on measures of team cognition. Together the results of these analyses address the overarching research question the current study investigates. Specifically, the first section addresses RQ1, while the second section focuses on RQ2. Statistical assumptions for tests were met for each analysis unless stated otherwise, and general acceptance of AI was implemented as a covariate when its effect on the data neared significance (p <.10) (Table 2).

4.1. Human task performance

A repeated-measures ANOVA was run to determine the effect of perceived teammate artificiality and task difficulty on human task performance in human-AI teams, directly addressing RQ1. The main effect of task difficulty on human task performance was found to be significant ($F(2,56) = 509.70, p < .001, \eta_p^2 = .95$). Holm corrected post-hoc tests on the main effect of task difficulty revealed that the effects of all three levels of task difficulty were significantly different from one another ($p < 0.001$). Specifically, participants performed best in easy difficulty tasks ($M = 108.58, SE = .493$), worst in hard difficulty tasks ($M = 130.39, SE = .493$), and in between for medium difficulty tasks ($M = 121.64$, $SE = .493$), which was expected and indicates a successful manipulation of task difficulty. The main effect of perceived teammate artificiality on human task performance was also found to be significant ($F(1,28) = 8.25, p < .01, \eta_p^2 = .23$), specifically, the the HHA condition ($M = 121.04, SE = .413$), performed significantly worse than the HHH condition overall ($M = 119.36, SE = .413$).

The interaction between perceived teammate artificiality and task difficulty was also found to be significant and was disordinal in nature ($F(2,56) = 37.58, p < .001, \eta_p^2 = .57$). Post-hoc analysis of the disordinal interaction effect between task difficulty and perceived teammate artificiality using Holm corrected tests indicated that the HHH condition had significantly better performance on the easy maps, followed by the medium maps and then hard maps ($p < 0.001$). This trend was also true for the HHA condition ($p < 0.01$), which also performed significantly worse than the HHH condition in each difficulty and comparison ($p < 0.01$), except for the hard maps ($p < 0.001$). The disordinal interaction occurred in the hard maps where the HHA condition ($M = 127.80$, $SD = 2.07$) significantly outperformed the HHH condition ($M = 132.98, SD = 3.30$) (Figure 2).

At a high level, these results conclude that when humans perceive that they are working with an artificial teammate, it will negatively impact their performance and that high difficulty tasks reduce their performance. However, additional analysis investigating builds upon the findings of RQ1 and determines that this is not always the case, leading to a more complex conclusion of the effects of perceived teammate artificiality on human task performance. The disordinal interaction effect is especially interesting regarding RQ1, as participant performance is worse when humans perceive that they are working with an artificial teammate up until they work on hard difficulty levels in the current task. This finding lends credence to the conclusion that humans are more receptive to agent teammates when faced with a difficult task, and this reception enables them to coordinate with their agent teammate more efficiently. Thus, the conclusions to RQ1 can be summed up by saying that humans' task performance is impacted by whether or not they perceive their teammate to be an agent; however, whether that impact is beneficial or harmful is based on the difficulty of the task they were completing with their agent teammate.

4.2. Team cognition

In addition to understanding the effects of perceived teammate artificiality and task difficulty on humans' task performance, this study also evaluates how it can impact the formation of team cognition (RQ2), an essential component of team development and performance. This analysis is broken up into three components of team cognition: (1) the convergence of mental models for team-related knowledge, (2) the convergence of mental models for task-related knowledge, and (3) the perceived team cognition participants reported. Finally, three exploratory mediation models were tested to assess the mediating effect of human task performance in varying task difficulties on the team's perceived team cognition levels. Each of these components individually provides a comprehensive analysis on the outcome of team cognition to thoroughly answer RQ2.

4.2.1. Team mental model similarity

For (1), a one-way ANCOVA was run to test the effect of perceived teammate artificiality on participants' team mental model convergence while controlling for participants general acceptance of AI ($p = 0.052$) (Figure 3). The ANCOVA neared significance ($F(1,27) = 3.89, p = .059, \eta_p^2 = .13$)

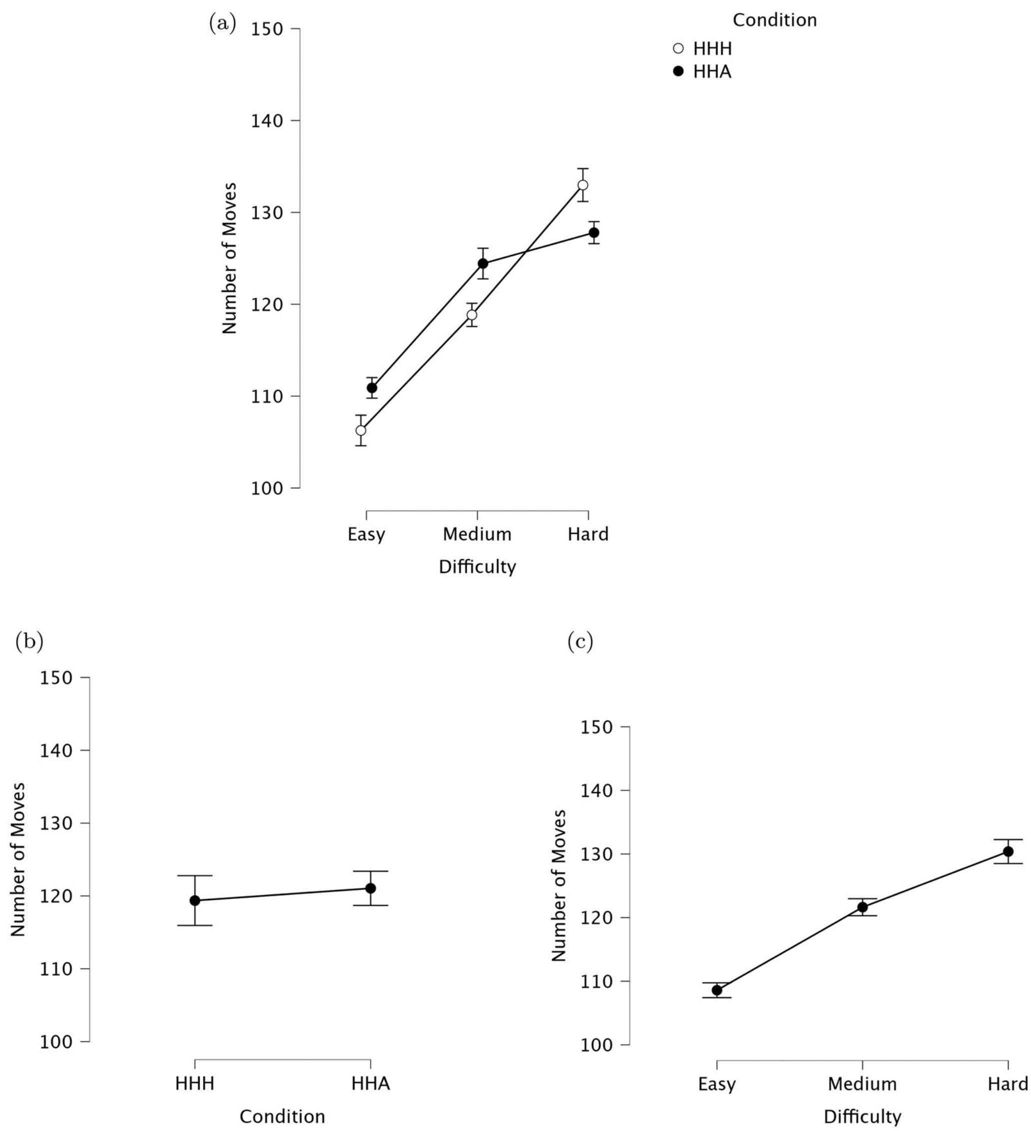

Figure 2. Measures of human task performance. (a) Human task performance for each team condition across easy, medium, and hard map difficulties. Please note that the higher the number of moves taken to collect the objectives the poorer the performance. Error bars denote 95% confidence intervals. (b) Overall human task performance for each team condition. Error bars denote 95% confidence intervals. (c) Overall human task performance for each team condition. Error bars denote 95% confidence intervals.

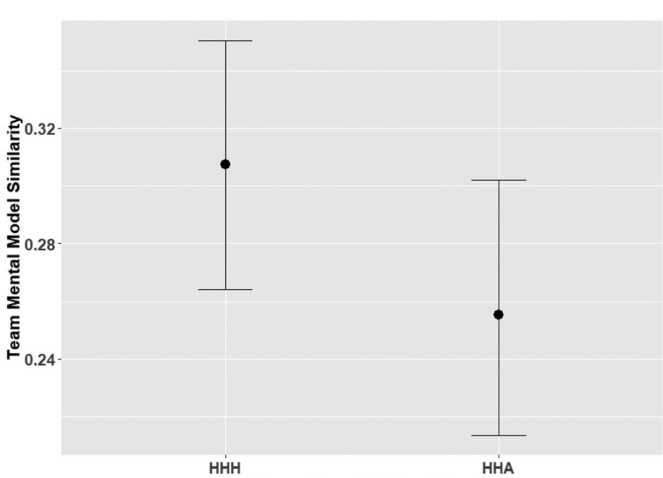

Figure 3. Team mental model similarity for both team conditions. Error bars denote 95% confidence intervals.

indicating that 13% of the variance in team mental model convergence could be explained by perceived teammate artificiality. Specifically, participants that perceived their third teammate as artificial trended towards having lower levels ($M = .255, SD = .09$) of team mental model convergence than participants that were told their teammate was human ($M = .26, SD = .09$). These results, while not significant, demonstrate that the knowledge of a teammate's artificiality may potentially serve as a roadblock that either prevents or slows down humans' ability to converge their team mental model with their other human teammates.

4.2.2. Task mental model similarity

For (2), an independent samples *t*-test was conducted to determine the effect of perceived teammate artificiality on the convergence of task mental models (Figure 4). Participants that were told their third teammate was artificial ($M = .52, SD = .13$) had higher task mental model convergence than teammates that were told they were working with a human ($M = .50, SD = .13$) but this difference was not statistically significant $t(28) = .42, p = .677, d = .15$. The 95% confidence interval for the mean difference was (−.12,.08). This insignificance is especially interesting given the nearly significant difference found in their team mental model similarity. This contrast in results may suggest that the roadblock created by one's knowledge of their teammate's artificiality may only impact the convergence of team-related mental models and that effect may not boil over into task-related shared knowledge.

4.2.3. Perceived team cognition

For (3), an independent samples t-test was conducted to assess the effect of perceived teammate artificiality on teams' perceived amount of team cognition. Participants that perceived their third teammate as artificial ($M = 10.13$, $SD = 4.77$) perceived lower (higher scores are worse for this measure) levels of team cognition than those that believed they were working with a human teammate ($M = 7.02$, $SD = 3.07$), and this difference was significant $t(28) = 2.12, p < .05, d = .78$. The 95% confidence interval for the mean difference was (−6.11, −.11).

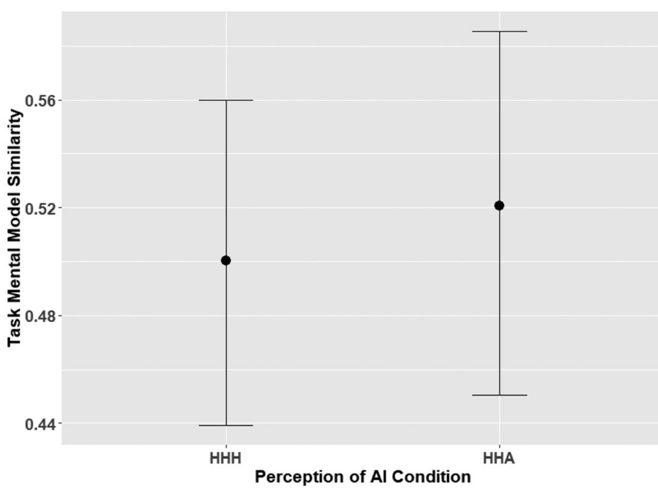

Figure 4. Task mental model similarity for both teaming conditions. Error bars denote 95% confidence intervals.

The experimental design also allowed for an analysis of perceived team cognition with human teammates and with AI teammates to be analyzed using an independent samples t-test. Due to significant heteroscedasticity between the two perceived teammate conditions, a Welch corrected t-test was utilized. The average perceived team cognition for AI teammates ($M = 14.33, SD = 8.12$) was greater (higher scores are worse) than that of human teammates ($M = 6.48$, $SD = 2.96$) and this difference was statistically different $t(15.88) = 3.63, p < .01, d = 1.29$. The 95% confidence interval for the mean difference was (−12.45, −3.26).

Finally, three mediation models were tested to investigate if the significant effect of perceived teammate artificiality on perceived team cognition was mediated by task difficulty, further answering RQ2 (Figure 5). This exploratory analysis was conducted based on the mediating effect of team dynamics on the relationship between perceived team cognition and team performance (Aubé et al., 2015), which the current study seeks to build upon by characterizing the influence of perceived teammate artificiality and task difficulty. Additionally, from a theoretical standpoint, the participants' perception of team cognition represents a subjective retrospective evaluation of their team's experience while performing the task. As such, the following mediation analyses model how human task performance in varying task difficulties mediates the significant relationship between the independent variable of perceived teammate artificiality and the participants' post-task perception of team cognition. Each mediation model was computed using the JASP mediation software (Love et al., 2019), which is based on the lavaan structural equation modeling package for R (Rosseel, 2012). Models were estimated using 5,000 maximum likelihood bootstrapped samples, delta method standard errors, and bias-corrected percentile

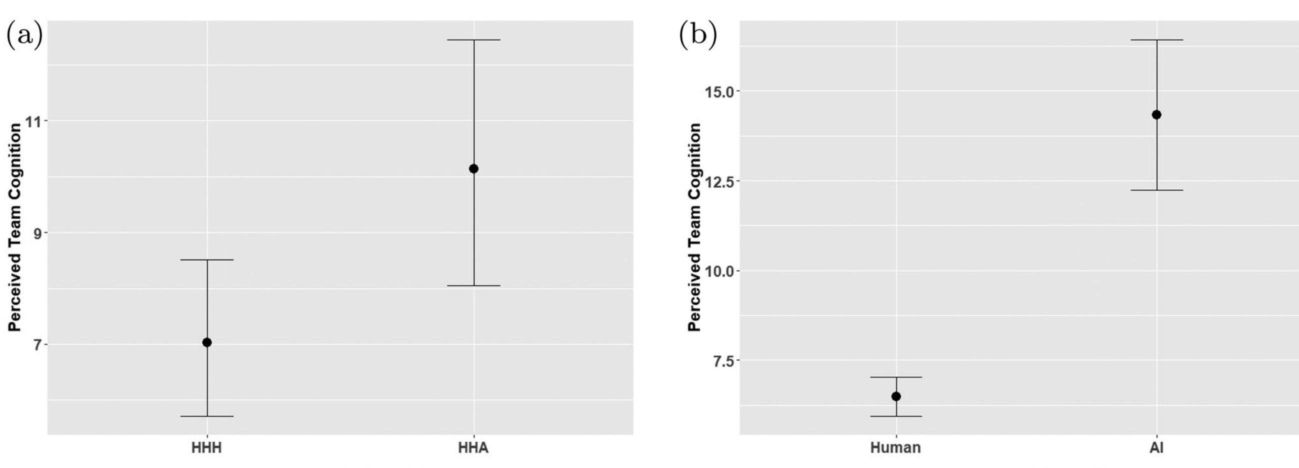

Figure 5. Measures of perceived team cognition. (a) Perceived team cognition for both team conditions. Error bars denote 95% confidence intervals. (b) Perceived team cognition for human and AI teammates. Error bars denote 95% confidence intervals.

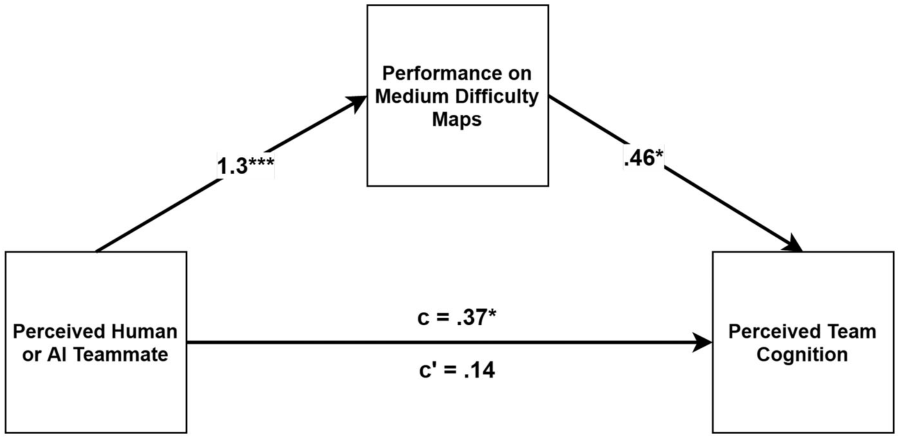

Figure 6. Mediating effect of medium map human task performance between perceived teammate artificiality and perceived team cognition. *Note.* Predictor variable is the categorical independent variable of perceived teammate artificiality, with a perceived human teammate coded as 0 and a perceived AI teammate coded as 1. Test of the indirect effect: ($z = 1.98, p < .05$).

bootstrapped confidence intervals as outlined by Biesanz and colleagues (Biesanz et al., 2010).

The first mediation model testing the indirect effect of human task performance on easy difficulty maps was found to be insignificant ($z = -1.32, p = .187$); however, the second mediation model testing the indirect effect of human task performance on medium difficulty maps was found to be significant ($z = 1.98, p < .05$). The effect of perceived teammate artificiality on perceived team cognition was significantly mediated by human task performance on maps of medium difficulty. As Figure 6 illustrates, the standardized parameter estimate between perceived teammate artificiality and perceived team cognition was significant (1.3) along with the standardized parameter estimate between human task performance on medium difficulty maps and perceived team cognition (.46). The indirect effect was $(1.3) * (.46) = .60$. The standardized indirect effects were computed for each of the 5,000 bootstrapped samples, and the 95% confidence interval was calculated using the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped standardized indirect effect was .60, and the 95% confidence interval had a minimum of .07 and a maximum of 1.5, which excluded zero and indicated that the indirect effect was significant ($p < .05$). Lastly, the mediation model analyzing the mediation effect of hard map human task performance was insignificant ($z = -0.347, p = .729$).

In response to RQ2, the results of the current analyses indicate that shared mental model similarity was not significantly influenced by perceived teammate artificiality. However, participants nonetheless believed that working with an AI teammate was detrimental to their team's shared understanding. While no definitive conclusions can be made regarding the apparent disparity between shared mental model similarity and perceived team cognition, it is possible that perceiving a teammate as artificial influences the various components of team cognition differently. Lastly, the interpretation of these exploratory mediation model analyses highlights that human-agent teams can suffer from the same mismatch between objective performance and subjective perceptions that human-only teams experience (Baugh & Graen, 1997). In the case of the current study, the significant mediating influence of human task performance on medium maps between perceived teammate condition and perceived team cognition contrasts with the significantly higher human task performance on hard maps for HHA teams. This mismatch suggests that human performance in tasks with the "just right" difficulty has the most significant influence over teams' perceived shared understanding. Alternatively, human task performance in easy tasks may be seen as too simple to be formative to team perceptions, while hard task difficulties may be too discouraging and or mentally demanding. Additionally, while the significant mediation effect above is technically a full mediation, it is important to state that true cases of total mediation are unlikely due to issues of power and the likely complexity of the true interaction between variables (Hayes, 2017).

In summary, the above results provide important and interesting insights into how humans' knowledge of whether or not their teammate is artificial can have direct impacts on a variety of critical teaming factors. First, answering RQ1 demonstrated that humans' actual task performance could be impacted by their knowledge of whether or not their teammate is human. A significant interaction effect also revealed that these impacts are not entirely constant and can fluctuate from harmful to beneficial based on the task's difficulty. Second, answering RQ2 demonstrates the consistently negative impacts that knowing a teammate is artificial can have on the development of team cognition; however, it is interesting to note that these effects are not consistent across different components of team cognition and are mediated by human task performance on specific task difficulties (namely medium). Overall, the above results provide fascinating insights into the field of human-agent teaming that have significant implications for how agent teammates should be incorporated into teams.

5. Discussion

The current study investigated the effects of perceived teammate artificiality on human task performance and cognition based on the difficulty of tasks. Specifically, we explored the differences in (1) human task performance and (2) team cognition between human-agent (HHA) and human-only teams depending on task difficulty. It is important to note that all of the purported AI agents in this research were all humans, which was critical to ensuring the measured effects were brought on only by the mere perception of the teammate. Thus, in the HHA condition, each human believed that they were one of two humans playing with the third AI teammate. In reality, all were humans akin to the participants in HHH condition as the confederate researcher portraying the AI teammate followed a script that mimicked human task performance (in place of expert-level AI agent performance (Foerster et al., 2018)). Furthermore, we did not allow communication between players to isolate the effect of implicit communication on team cognition and human-agent teams, and ensure participants, especially those within the HHA condition, did not suspect the AI teammate as a human (Entin & Serfaty, 1999; Hanna & Richards, 2014, 2015; Shively et al., 2017; Young et al., 2018).

The differences observed between the two conditions were a direct consequence of the perception of working with AI teammates rather than limited capabilities or different behavioral patterns used by AI, as outlined previously. Specifically, RQ1 found that the perception of working with an AI teammate negatively influences human task performance in human-AI teams. However, this effect was moderated by task difficulty such that hard difficulty tasks reversed the trend, and the perception of an AI teammate significantly improved human task performance. RQ2 found no significant effect of perceived teammate artificiality on shared mental model similarity but did find a significant difference between the two perceived team conditions in perceived team cognition. Teams that perceived their third teammate as an AI had significantly less perceived team cognition, and, overall, participants perceived more team cognition with human teammates than the perceived AI teammate. Additionally, human task performance in medium-difficulty tasks was the most influential factor in determining how participants rated their teams perceived team cognition.

5.1. Utilizing AI teammates to overcome difficult tasks

AI teammates bring several challenges to the table for human teammates whenever they operate in human-AI teams, with major limitations to communication ability and their potential to develop shared understanding (Cooke et al., 2013; Young et al., 2018). Humans expect AI teammates to have at least a basic ability to engage in these human-like behaviors (Zhang et al., 2021), if not excel at them. However, human teammates know that AI teammates are not currently capable of effectively engaging in these behaviors and must therefore adapt their own behavioral and communication patterns to accommodate these differences.

The current study's results indicate that even when behavioral and communication limitations are taken away, and all teammates are placed on a level playing field, human task performance is still significantly impacted by the simple perception that they were working with an AI teammate. This finding is significant because a behavior is emerging from the human teammates' perception that they are working with an AI agent, and it has a meaningful effect on their performance. Prior research has hinted that this could be due to differences in perception, communication, and trust in the capabilities of AI agents (Demir et al., 2018; McNeese et al., 2021; Musick et al., 2021; Walliser et al., 2017). Specifically, Demir and colleagues found that believing a teammate is artificial makes it more difficult for human-agent teams to plan and effectively anticipate their teammates' needs (Demir et al., 2018). This negative effect is partially because humans crave more social interaction when working in teams, which unfortunately AI teammates are not able to adequately facilitate (Walliser et al., 2017). However, the current study does not include any of these variables, indicating that the effect goes beyond negatively affecting team processes and exists at a fundamental level within the cognition of humans participating in human-AI teams.

Interestingly, this perceptual effect on human behavior was reversed when the difficulty of the task became hard and is not entirely expected as a recent review of human-autonomy teaming found that such teams suffered when exposed to greater task difficulties compared to human-human teams (O'Neill et al., 2020). A likely explanation for this divergent finding is that humans and AI teammates were placed on a level playing field in terms of teammate abilities (i.e., communication), which negated the objective disadvantages of AI teammates and isolated the perceptual effects instead. This finding allows for a better interpretation of the results as it allows the effect to be studied in varying task contexts. Specifically, this reversal could be a result of the human team members with a perceived AI teammate having more confidence than the perceived human-only team purely *because* they thought they had an AI teammate. Human teammates may have had more confidence in their team's ability to effectively perform in the hard-difficulty tasks because they believed their AI teammate would help their team in a way typical human teammates may not have been able to based on its automatic and expert level capabilities (de Visser & Parasuraman, 2011; Foerster et al., 2018). This finding aligns with the trend that autonomous agents can benefit teams under challenging environments or those that contain a great deal of uncertainty (Demir et al., 2015; McKendrick et al., 2014; de Visser & Parasuraman, 2011). With the supervision of unmanned air vehicles (UAV), McKendrick and colleagues (McKendrick et al., 2014) found that autonomous agents or decision aids can improve team performance under high task load (i.e., hard task difficulty). This finding is partly due to improved working memory capacities of human-agent teams, with the autonomous agent being able to harbor a higher working memory capacity than humans and ultimately direct and use it to execute complex tasks. With this study, we build on this prior work, with the finding that, specifically, that task difficulty should be considered when choosing to implement a traditional human-human team or a human-AI team. Additionally, the benefits to performance noted in the aforementioned studies can somewhat extend to human-AI teams as a placebo since the human-AI teams in the current study did not necessarily benefit from the decreased workload and increased working memory offered by real expert level agents in those previous studies.

Based on the current findings, it appears that the general prior knowledge of AI teammates' high taskwork abilities alone is enough of a placebo to enhance human task performance in specific high difficulty tasks. This finding extends previous literature by showcasing that participants' that simply believe their team is diverse in AI teammate membership have an apparent advantage in high difficulty tasks over teams that believe their membership is homogeneous. This trend is seen in human-human teams where diverse (i.e., personality and gender) heterogeneous teams outperform homogeneous teams in high difficulty tasks (Bowers et al., 2000). Overall, in our work, we show that whether or not human-AI teams are better than human-only teams can depend on the difficulty of the task (e.g., using robots for hazardous environments). This finding also serves as an example of leveraging the known advantages offered by AI teammates instead of focusing on their aforementioned shortcomings; however, a deeper exploration is needed to clearly understand what tasks humans perceive as difficult to better inform the practicality of a human-agent team-up. For example, humans may find a variety of optimization problems in supply chain management exceedingly difficult to tackle without help from an AI teammate, while more creative tasks like conceptualizing the consequences of a particular manufacturing layout on employee morale and efficiency are still complex but manageable for experienced employees. It is likely that task difficulty, especially hard tasks, exist on a gradient and can range from nearly impossible without an AI teammate, to more reasonable hard difficulty tasks like those mentioned previously. Future research should investigate humans perceptions of hard difficulty tasks to better understand how humans perceive such tasks and whether the effects noted in the current study change based on those perceptions.

5.2. Accounting for the potentially negative impacts of AI teammates on team dynamics and team cognition

The current study also investigated the influence of perceived teammate artificiality on team cognition, which produced mixed results. Specifically, while no significant effect of perceived teammate artificiality on shared mental model similarity was found, the perceived level of team cognition was significantly affected. This finding continues to add to the growing body of evidence that humans find it difficult to perceive AI as full teammates despite their rapidly expanding abilities. It would seem their continued deficiencies in developing shared understanding, communication, and human-like social behavior serve as extremely limiting factors. These specific deficiencies are highlighted based on recent research that found humans frequently teaming up with AI list these characteristics as their most desired from AI when working together with them (Zhang et al., 2021). These deficiencies are also highly related to the development and usage of team cognition in human-AI teams. Specifically, Liang and colleagues emphasize the fact that teams need to communicate in order to establish and maintain common ground for the purpose of agreeing on joint intentions and shared goals (Liang et al., 2019, p.10). Otherwise, human team members tend to silo the AI based on the perception that communication (implicit or explicit) with the agent is pointless and unpredictable. As a result, they tend to leave the agent out of teamwork which ultimately affects team performance (Liang et al., 2019; Musick et al., 2021). Though increasing the accountability pressures of human teammates has been shown to influence resource exchange strategies in human-AI teams such that humans shared more with their AI teammate (León et al., 2021). The negative findings from the current study may also extend to HAA teams, which differ from the HHA teams in the current study by having a team composition that is primarily AI instead of human. A single human working alone with multiple AI teammates may experience these same effects but exacerbated as their typical idea of teaming and the processes they have come to associate with it are thrown out of the window almost entirely. Additionally, if the negative effects are magnified the improvement in performance for hard difficulty tasks may be negated in these teams. Further research is necessary to explore how a sole human in a human-AI team, whether it is a dyad (human-AI) or larger (HAA), is affected by the change in typical teaming functions.

Taking steps towards overcoming the adverse effects of perception likely lies in leveraging the presentation of the AI, the training the humans and AI receive together and focusing on the unique capabilities that AI teammates *do* bring to the table that can offset their current limitations. These advantages include their expert-level task ability, which can help serve as an example of effective taskwork for human teammates, and their ability to store and manage vast amounts of information all at once while simultaneously performing other tasks. Again, these advantages become clearer in varying task contexts and difficulties, as was shown in the current study's findings. However, not to be overlooked is the finding that human task performance in medium difficulty tasks was a significant mediator for participants' perception of the team's level of team cognition. This finding extends existing literature connecting team dynamics as a mediating factor between perceived team cognition and human task performance (Aubé et al., 2015). Specifically, by including perceived teammate artificiality in the category of team dynamics, the current study finds that human task performance in medium difficulty tasks mediates the relationship between perceived teammate artificiality and perceived levels of team cognition.

One interpretation of these results could be that easy difficulty tasks are too easy to significantly shape participants' perception of their team's shared understanding. Alternatively, the lack of a mediating effect of hard difficulty maps may indicate that these maps either came too late to influence participants' judgment of shared understanding or that hard-difficulty maps were too discouraging and took too much cognitive load. From a practical standpoint, these results should encourage enhanced training programs for human-AI teams, especially for practitioners that seek to maximize their teams' level of perceived team cognition, which is related to eventual team performance (Aubé et al., 2015). By providing positive experiences with AI teammates in training to encourage increased levels of trust (Hafizoglu & Sen, 2018a, 2018b), and using medium difficulty tasks to enhance perceived team cognition, human-AI teams can become more effective when they are functionally deployed.

5.3. Limitations and future research

The current study has some limitations that should be considered when interpreting these results for future research or application onto practicing human-AI teams. The first limitation is the experimental design, which required all participants to not engage in any explicit communication. While this design is necessary to isolate the effect of perceiving a teammate as artificial, it partially limits some of the results' practical applicability. Another limitation to keep in mind is the task itself, which was digital and relatively simple to examine basic effects through high levels of experimental control. As such, these findings should be explored in additional experimental task contexts and environments realistic to real-world human-AI teaming. Future research should also focus on uncovering additional variables that influence the relationship between perceptions of AI teammates and eventual human task performance. For example, individual differences likely play a prominent role in determining just how strongly the perception of AI teammates affects team cognition and individual human task performance. While the current study investigated the effect of perceived teammate artificiality on teams of three, these research questions can also apply to dyads. Dyads are a critical area of human-AI teaming and interaction research as one on one partnerships are becoming increasingly common within applied contexts like warehouses and factories (Wilson & Daugherty, 2018). Studying the nature of these effects on team outcomes within dyads and other varied team sizes and compositions should be a focus of future research. Dyads would also not be affected by the limitation on communication between teammates outlined here in the current study. As such, research with dyads could elicit results that would be highly applicable to practical contexts and even advance research on human-AI communication specifically.

6. Conclusion

The current study investigated two research questions focusing on how the perception of working with an AI teammate influenced individual human task performance and team cognition. The study found that the simple perception of working with an AI negatively influenced human task performance overall but improved it in hard-difficulty tasks. There was no significant effect of perceived teammate artificiality on shared mental model similarity. However, perceptions of team cognition were negatively affected. The development of perceived team cognition was also influenced differently based on participants' task performance in the three task difficulties. Specifically, human task performance in medium difficulty tasks significantly mediated the relationship between perceived teammate artificiality and participants' perceived level of team cognition, while easy and hard difficulty tasks did not. The results of this study demonstrate how AI teammates may not only directly benefit individual task performance in hard tasks due to their computational abilities, but those direct impacts may further mediate and improve critical perceptions in human-AI teams. The results of this artificial will help guide the creation and implementation of AI teammates alongside humans to ensure humans benefit from their computational abilities.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The Office of Naval Research (award number N000142112336) and the Air Force Office of Scientific Research (award number FA9550-20-1-0342) were the funders for this project.

ORCID

Beau G. Schelble http://orcid.org/0000-0003-3704-697X
Christopher Flathmann http://orcid.org/0000-0002-5448-2610
Nathan J. McNeese http://orcid.org/0000-0002-9143-2460
Thomas O'Neill http://orcid.org/0000-0002-8336-5763
Richard Pak http://orcid.org/0000-0001-9145-6991
Moses Namara http://orcid.org/0000-0002-0099-8275

References

Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. *Cities*, *89*, 80–91. https://doi.org/10.1016/j.cities.2019.01.032

Amon, M. J., Vrzakova, H., & D'Mello, S. K. (2019). Beyond dyadic coordination: Multimodal behavioral irregularity in triads predicts facets of collaborative problem solving. *Cognitive Science*, *43*(10), e12787. https://doi.org/10.1111/cogs.12787

Aubé, C., Rousseau, V., & Tremblay, S. (2015). Perceived shared understanding in teams: The motivational effect of being 'on the same page'. *British Journal of Psychology (London, England: 1953)*, *106*(3), 468–486. https://doi.org/10.1111/bjop.12099

Aubert, M. C., Bader, H., & Hauser, K. (2018). Designing multimodal intent communication strategies for conflict avoidance in industrial human-robot teams. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1018–1025). IEEE. https://doi.org/10.1109/ROMAN.2018.8525557

Baugh, S. G., & Graen, G. B. (1997). Effects of team gender and racial composition on perceptions of team performance in cross-functional teams. *Group & Organization Management*, *22*(3), 366–383. https://doi.org/10.1177/1059601197223004

Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*, *45*(4), 661–701. https://doi.org/10.1080/00273171.2010.498292

Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in work teams: A meta-analysis. *Small Group Research*, *31*(3), 305–327. https://doi.org/10.1177/104649640003100303

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, *39*(3/4), 324–345. https://doi.org/10.2307/2334029

Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, *37*(8), 729–758. https://doi.org/10.1080/10447318.2020.1841438

Chen, J. Y., & Barnes, M. J. (2012). Supervisory control of multiple robots: Effects of imperfect automation and individual differences. *Human Factors*, *54*(2), 157–174. https://doi.org/10.1177/0018720811435843

Chen, J. Y., Barnes, M. J., & Harper-Sciarini, M. (2011). Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *41*(4), 435–454. https://doi.org/10.1109/TSMCC.2010.2056682

Chen, J. Y., Barnes, M. J., Wright, J. L., Stowers, K., & Lakhmani, S. G. (2017). Situation awareness-based agent transparency for human-autonomy teaming effectiveness. In *Micro-and nanotechnology sensors, systems, and applications IX* (Vol. 10194, p. 101941V). International Society for Optics and Photonics. https://doi.org/10.1117/12.2263194

Converse, S., Cannon-Bowers, J. A., & Salas, E. (1993). *Shared mental models in expert team decision making*, *221*, 221–246. Lawrence Erlbaum Associates, Inc., New Jersey.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–285. https://doi.org/10.1111/cogs.12009

Cooke, N. J., Kiekel, P. A., Salas, E., Stout, R., Bowers, C., & Cannon-Bowers, J. (2003). Measuring team knowledge: A window to the cognitive underpinnings of team performance. *Group Dynamics: Theory, Research, and Practice*, *7*(3), 179–199. https://doi.org/10.1037/1089-2699.7.3.179

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. J. (2000). Measuring team knowledge. *Human Factors*, *42*(1), 151–173. https://doi.org/10.1518/001872000779656561

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, *50*(6), 456–473. https://doi.org/10.1037/h0057173

De Visser, E., & Parasuraman, R. (2011). Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload. *Journal of Cognitive Engineering and Decision Making*, *5*(2), 209–231. https://doi.org/10.1177/1555343411410160

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, *14*(1), 1–14. https://doi.org/10.1037/a0017455

Demir, M., McNeese, N. J., & Cooke, N. J. (2018). The impact of perceived autonomous agents on dynamic team behaviors. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(4), 258–267. https://doi.org/10.1109/TETCI.2018.2829985

Demir, M., McNeese, N. J., Cooke, N. J., Ball, J. T., Myers, C., & Frieman, M. (2015). Synthetic teammate communication and coordination with humans. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 951–955. https://doi.org/10.1177/1541931215591275

Dorton, S. L., Hall, R. A. (2021). Collaborative human-ai sensemaking for intelligence analysis. In International *conference on human-computer interaction* (pp. 185–201). Springer: Cham, Switerland. https://doi.org/10.1007/978-3-030-77772-2_12

Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human Factors*, *41*(2), 312–325. https://doi.org/10.1518/001872099779591196

Fan, X., & Yen, J. (2010). Modeling cognitive loads for evolving shared mental models in human–agent collaboration. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *41*(2), 354–367. https://doi.org/10.1109/TSMCB.2010.2053705

Fiore, S. M., Salas, E., & Cannon-Bowers, J. A. (2001). *Group dynamics and shared mental model development* (Vol. 234). Lawrence Earlbaum Associates Publishers.

Foerster, J., Chen, R. Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., & Mordatch, I. (2018). Learning with opponent-learning awareness. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems* (pp. 122–130).

Gorman, J. C., & Cooke, N. J. (2011). Changes in team cognition after a retention interval: The benefits of mixing it up. *Journal of Experimental Psychology. Applied*, *17*(4), 303–319. https://doi.org/10.1037/a0025149

Gorman, J. C., Cooke, N. J., Pedersen, H. K., Winner, J., Andrews, D., & Amazeen, P. G. (2006). Changes in team composition after a break: Building adapative command-and-control teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 487–491. https://doi.org/10.1177/154193120605000358

Hafizoglu, F. M., & Sen, S. (2018a). The effects of past experience on trust in repeated human-agent teamwork. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 514–522). International Foundation for Autonomous Agents and Multiagent Systems.

Hafizoglu, F. M., & Sen, S. (2018b). Reputation based trust in human-agent teamwork without explicit coordination. In *Proceedings of the 6th international conference on human-agent interaction* (pp. 238–245). Association for Computing Machinery, New York, NY, USA.

Hanna, N., & Richards, D. (2014). The impact of communication on a human-agent shared mental model and team performance. In *Proceedings of the 2014 international conference on autonomous agents and multi-agent systems* (pp. 1485–1486). International Foundation for Autonomous Agents and Multiagent Systems. Richland, SC.

Hanna, N., & Richards, D. (2015). The impact of virtual agent personality on a shared mental model with humans during collaboration. In *Proceedings of the 2015 international conference on autonomous agents and multiagent systems* (pp. 1777–1778). International Foundation for Autonomous Agents and Multiagent Systems. Richland, SC.

Hanna, N., & Richards, D. (2018). The impact of multimodal communication on a shared mental model, trust, and commitment in human–intelligent virtual agent teams. *Multimodal Technologies and Interaction*, *2*(3), 48. https://doi.org/10.3390/mti2030048

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Kaminka, G. A. (2013). Curing robot autism: A challenge. In *Proceedings of the 2013 international conference on autonomous agents and multi-agent systems* (pp. 801–804). International Foundation for Autonomous Agents and Multiagent Systems. Richland, SC.

Kelley, J. F. (1983). An empirical methodology for writing user-friendly natural language computer applications. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 193–196). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/800045.801609

Kelley, J. F. (2018). Wizard of oz (WoZ) a yellow brick journey. *Journal of Usability Studies*, *13*(3), 119–124.

Kepuska, V., & Bohouta, G. (2018). Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In *2018 IEEE 8th Annual Computing and Communication*

Workshop and Conference (CCWC) (pp. 99–103). IEEE. https://doi.org/10.1109/CCWC.2018.8301638

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, *20*(2), 403–437. https://doi.org/10.1177/014920639402000206

Langan-Fox, J., Code, S., & Langfield-Smith, K. (2000). Team mental models: Techniques, methods, and analytic approaches. *Human Factors*, *42*(2), 242–271. https://doi.org/10.1518/001872000779656534

Lee, M., & Johnson, T. E. (2008). Understanding the effects of team cognition associated with complex engineering tasks: Dynamics of shared mental models, task-SMM, and team-SMM. *Performance Improvement Quarterly*, *21*(3), 73–95. https://doi.org/10.1002/piq.20032

León, G. A., Chiou, E. K., & Wilkins, A. (2021). Accountability increases resource sharing: Effects of accountability on human and AI system performance. *International Journal of Human–Computer Interaction*, *37*(5), 434–444. https://doi.org/10.1080/10447318.2020.1824695

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*(2), 273–307. https://doi.org/10.1111/j.1744-6570.2008.00114.x

Liang, C., Proft, J., Andersen, E., & Knepper, R. A. (2019). Implicit communication of actionable information in human-AI teams. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery. New York, NY, USA.

Lim, B.-C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403–418. https://doi.org/10.1002/job.387

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., Ly, A., Gronau, Q. F., Smíra, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E.-J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*(2), 1–17. https://doi.org/10.18637/jss.v088.i02

Mabkhot, M. M., Al-Ahmari, A. M., Salah, B., & Alkhalefah, H. (2018). Requirements of the smart factory system: A survey and perspective. *Machines*, *6*(2), 23. https://doi.org/10.3390/machines6020023

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review*, *26*(3), 356–376. https://doi.org/10.2307/259182

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Cannon-Bowers, J. A., & Salas, E. (2005). Scaling the quality of teammates' mental models: Equifinality and normative comparisons. *Journal of Organizational Behavior*, *26*(1), 37–56. https://doi.org/10.1002/job.296

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. https://doi.org/10.1037/0021-9010.85.2.273

McKendrick, R., Shaw, T., de Visser, E., Saqer, H., Kidwell, B., & Parasuraman, R. (2014). Team performance in networked supervisory control of unmanned air vehicles: Effects of automation, working memory, and communication content. *Human Factors*, *56*(3), 463–475. https://doi.org/10.1177/0018720813496269

McNeese, N. J., Demir, M., Chiou, E. K., & Cooke, N. J. (2021). Trust and team performance in human–autonomy teaming. *International Journal of Electronic Commerce*, *25*(1), 51–72. https://doi.org/10.1080/10864415.2021.1846854

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. https://doi.org/10.1177/0018720817743223

McNeese, N. J., Schelble, B. G., Canonico, L. B., & Demir, M. (2021). Who/what is my teammate? team composition considerations in human–ai teaming. *IEEE Transactions on Human-Machine Systems*, *51*(4), 288–299. https://doi.org/10.1109/THMS.2021.3086018

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for multi-uxv management. *Human Factors*, *58*(3), 401–415. https://doi.org/10.1177/0018720815621206

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management*, *36*(4), 876–910. https://doi.org/10.1177/0149206309356804

Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods*, *3*(2), 123–165. https://doi.org/10.1177/109442810032001

Moreland, R. L. (2010). Are dyads really groups? *Small Group Research*, *41*(2), 251–267. https://doi.org/10.1177/1046496409358618

Musick, G., O'Neill, T. A., Schelble, B. G., McNeese, N. J., & Henke, J. B. (2021). What happens when humans believe their teammate is an AI? An investigation into humans teaming with autonomy. *Computers in Human Behavior*, *122*, 106852. https://doi.org/10.1016/j.chb.2021.106852

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2020). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors*, https://doi.org/10.1177/0018720820960865

Pape, L. J. (1998). *The effect of personality characteristics on team member schema similarity* [Phd thesis]. Wright State University

Pontiggia, A., & Virili, F. (2010). Network effects in technology acceptance: Laboratory experimental evidence. *International Journal of Information Management*, *30*(1), 68–77. https://doi.org/10.1016/j.ijinfomgt.2009.07.001

Rentsch, J., McNeese, M. D., Pape, L. J., Burnett, D. D., & Menard, D. M. (1998). *Testing the effects of team processes on team member schema similarity and team performance: Examination of the team member schema similarity model.* National Technical Information Service, Springfield, Virginia.

Resick, C. J., Dickson, M. W., Mitchelson, J. K., Allison, L. K., & Clark, M. A. (2010). Team composition, cognition, and effectiveness: Examining mental model similarity and accuracy. *Group Dynamics: Theory, Research, and Practice*, *14*(2), 174–191. https://doi.org/10.1037/a0018444

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (beta). *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Saavedra, R., Earley, P. C., & Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology*, *78*(1), 61–72. https://doi.org/10.1037/0021-9010.78.1.61

Santos, C. M., Uitdewilligen, S., & Passos, A. M. (2015). A temporal common ground for learning: The moderating effect of shared mental models on the relation between team learning behaviours and performance improvement. *European Journal of Work and Organizational Psychology*, *24*(5), 710–725. https://doi.org/10.1080/1359432X.2015.1049158

Schaarschmidt, M., Kuhnle, A., & Fricke, K. (2017). *TensorForce: A TensorFlow library for applied reinforcement learning.* https://github.com/tensorforce/tensorforce

Schelble, B. G., Flathmann, C., & McNeese, N. (2020). Towards meaningfully integrating human-autonomy teaming in applied settings. In *Proceedings of the 8th international conference on human-agent interaction* (pp. 149–156). Association for Computing Machinery. New York, NY, United States. https://doi.org/10.1145/3406499.3415077

Schelble, B. G., Flathmann, C., McNeese, N., Freeman, G., & Mallick, R. (2022). Let's think together! Assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction*, *6*(GROUP), 1–29. https://doi.org/10.1145/3492832

Schmidt, C., Kude, T., Heinzl, A., & Mithas, S. (2014). *How agile practices influence the performance of software development teams: The role of shared mental models and backup.* https://aisel.aisnet.org/icis2014/proceedings/ProjectManagement/15

Schvaneveldt, R. W. (1990). *Pathfinder associative networks: Studies in knowledge organization.* Ablex Publishing.

Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V., Johnson, W. W. (2017). Why Human-Autonomy Teaming? In *International conference on applied human factors and ergonomics* (pp. 3–11). Springer.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Spotts, H. E., & Chelte, A. F. (2005). Evaluating the effects of team composition and performance environment on team performance. *Journal of Behavioral and Applied Management*, *6*(2), 127–140. https://doi.org/10.21818/001c.14526

Stephanidis, C., Salvendy, G., Antona, M., Chen, J. Y. C., Dong, J., Duffy, V. G., Fang, X., Fidopiastis, C., Fragomeni, G., Fu, L. P., Guo, Y., Harris, D., Ioannou, A., Jeong, K-a Kate)., Konomi, S., Krömker, H., Kurosu, M., Lewis, J. R., Marcus, A., … Zhou, J. (2019). Seven HCI grand challenges. *International Journal of Human–Computer Interaction*, *35*(14), 1229–1269. https://doi.org/10.1080/10447318.2019.1619259

Stout, R. J., Cannon-Bowers, J. A., & Salas, E. (2017). The role of shared mental models in developing team situational awareness: Implications for training. In *Situational awareness* (pp. 287–318). Routledge.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *41*(1), 61–71. https://doi.org/10.1518/001872099779577273

Tokadl i, G., Dorneich, M. C., & Matessa, M. (2021). Evaluation of playbook delegation approach in human-autonomy teaming for single pilot operations. *International Journal of Human–Computer Interaction*, *37*(7), 703–716. https://doi.org/10.1080/10447318.2021.1890485

Van Riemsdijk, M. B., Jonker, C. M., Lesser, V. (2015). Creating socially adaptive electronic partners: Interaction, reasoning and ethical challenges. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1201–1206).

Ven, A. H. V. d., Delbecq, A. L., & Koenig, R. Jr. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, *41*(2), 322–338. https://doi.org/10.2307/2094477

Walliser, J. C., Mead, P. R., & Shaw, T. H. (2017). The perception of teamwork with an autonomous agent enhances affect and performance outcomes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 231–235. https://doi.org/10.1177/1541931213601541

Wehbe, R. R., Lank, E., & Nacke, L. E. (2017). Left them 4 dead: Perception of humans versus non-player character teammates in cooperative gameplay. In *Proceedings of the 2017 conference on designing interactive systems* (pp. 403–415). Association for Computing Machinery, New York, NY, United States.

Williams, K. D. (2010). Dyads can be groups (and often are). *Small Group Research*, *41*(2), 268–274. https://doi.org/10.1177/1046496409358619

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative Intelligence: Humans and AI Are Joining Forces, *96*(4), 114–123. Harvard Business Review.

Wright, M. C., & Kaber, D. B. (2005). Effects of automation of information-processing functions on teamwork. *Human Factors*, *47*(1), 50–66. https://doi.org/10.1518/0018720053653776

Yokoi, R., Eguchi, Y., Fujita, T., & Nakayachi, K. (2021). Artificial intelligence is trusted less than a doctor in medical treatment decisions: Influence of perceived care and value similarity. *International Journal of Human–Computer Interaction*, *37*(10), 981–990. https://doi.org/10.1080/10447318.2020.1861763

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738

Zhang, R., McNeese, N. J., Freeman, G., & Musick, G. (2021). An ideal human. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 1–25. https://doi.org/10.1145/3432945

About the Authors

Beau Schelble is a Ph.D. student at Clemson University studying Human-Centered Computing in the Team Research in Computational Environments (TRACE) Research Group within the School of Computing. His research interests lie in human-AI teaming, team cognition, and applications of human-AI partnerships.

Christopher Flathmann is a Ph.D. student at Clemson University studying Human-Centered Computing in the Team Research in Computational Environments (TRACE) Research Group within the School of Computing. His research interests lie in acceptance of AI technologies, human-AI teaming, and human-centered AI.

Nathan McNeese is the College of Engineering, Computing and Applied Sciences (CECAS) Dean's Professor and Assistant Professor of Human-Centered Computing within the School of Computing at Clemson University. He directs the TRACE lab and received his Ph.D. in Information Science in 2014 from Pennsylvania State University.

Tom A. O'Neill is the Director of the Individual and Team Performance Lab and Associate Professor of Industrial-Organizational Psychology at the University of Calgary, Canada. His research considers high-performance teamwork, virtual team effectiveness, conflict and conflict management, personality, and assessment.

Richard Pak is currently a Professor in the Department of Psychology at Clemson University. He received his Ph.D. in Engineering Psychology in 2005 from the Georgia Institute of Technology.

Moses Namara is currently a privacy researcher at Meta. He received his Ph.D. in Human-Centered Computing in 2022 from Clemson University.