Computers in Human Behavior 165 (2025) 108560

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Understanding the processes of trust and distrust contagion in Human–AI Teams: A qualitative approach

Wen Duan [a,*], Shiwen Zhou [b], Matthew J. Scalia [b], Guo Freeman [a], Jamie Gorman [b], Michael Tolston [c], Nathan J. McNeese [a], Gregory Funke [c]

[a] Human-Centered Computing, Clemson University, McAdams Hall, 821 McMillan Rd, Clemson, SC 29631, USA
[b] Human Systems Engineering; Center for Human, Artificial Intelligence, and Robot Teaming (CHART), Arizona State University, 7418 E. Innovation Way South Mesa, AZ 85212, USA
[c] Air Force Research Laboratory Wright-Patterson Air Force Base, OH 45433, USA

ARTICLE INFO

Keywords:
Human–AI teaming
Trust contagion
Qualitative method

ABSTRACT

The success of human–AI teams (HATs) requires humans to work with AI teammates in trustful ways. However, trust does not exist in a vacuum but forms through and can be influenced by interactions among teammates, leading to understudied questions about how trust or distrust can be spread within a HAT. Drawing on interviews with 36 participants who collaborated in a three-member human–AI team, we explore human perceptions of and reactions to a human or AI teammate's (dis)trust spread about an AI teammate, and uncover the process and impact of such spread. Our findings highlight that a trustworthy (dis)trust spreader can catalyze trust contagion within a human–AI team through various social and cognitive processes. We provide one of the first empirical investigations into specific ways through which trust or distrust can be spread within HATs and people's perceptions of such spread. We thus contribute to the effective design of AI teammates and human–AI team dynamics that foster an appropriate level of trust in future HATs.

1. Introduction

Trust in human–AI teaming has been a growing research agenda in HCI (human–computer interaction) and CSCW (computer-supported cooperative work) communities (Duan et al., 2024; Flathmann, Duan, Mcneese, Hauptman, & Zhang, 2024; Hauptman, Schelble, Duan, Flathmann, & McNeese, 2024; Zhang, Chong, Kotovsky, & Cagan, 2023; Zhang et al., 2024), as trust plays a critical role in both effective teamwork (Costa, Fulmer, & Anderson, 2018; Mayer, Davis, & Schoorman, 1995; McAllister, 1995) and effective use of AI technology (Bansal et al., 2022; Glikson & Woolley, 2020). Indeed, decades of team and organizational research has suggested that teams characterized by strong mutual trust typically outperform and operate more efficiently compared to teams marked by a lack of trust (Breuer, Hüffmeier, & Hertel, 2016; De Jong, Dirks, & Gillespie, 2016). Trust among team members fosters better cooperation (Balliet & Van Lange, 2013), boosts team satisfaction (Chou, Wang, Wang, Huang, & Cheng, 2008), facilitates knowledge exchange (Szulanski, Cappetta, & Jensen, 2004), and exerts positive influences on various other critical factors essential for enhanced team outcomes (Costa et al., 2018). As AI technology becomes more integrated into human teams (Schelble, Flathmann, McNeese, Freeman, & Mallick, 2022; Zhang, Lee, & Carter, 2022; Zhang, McNeese, Freeman, & Musick, 2021), the extent to which humans and AI can work together in trustful ways has the potential to significantly determine the success and effectiveness of human–AI teams (HATs) (McNeese, Demir, Chiou, & Cooke, 2021).

However, while there is extensive research on how humans' trust impacts their adoption (Jacovi, Marasović, Miller, & Goldberg, 2021) and effective use of AI (Bansal et al., 2022; Lee & See, 2004; Lu & Yin, 2021), as well as the importance of AI's reliability (Avril, 2023; Rieger, Roesler, & Manzey, 2022), transparency (Bobko et al., 2023; Chen et al., 2016; Ehsan, Liao, Muller, Riedl, & Weisz, 2021), and explainability (Ehsan, Saha, De Choudhury, & Riedl, 2023; Wang, Pynadath, & Hill, 2016) in increasing and calibrating humans' trust, little is known regarding how humans' trust in AI teammates actually forms, evolves, and changes during **team** interactions, especially in response to team-related factors (Costa et al., 2018; Ulfert, 2020) such as fellow teammates' opinions (Grosser, Kidwell, & Labianca, 2012; Spoelma & Hetrick, 2021; Van de Bunt, Wittek, & de Klepper, 2005). Gaining such an understanding is critical as trust does not exist in a vacuum. Rather, research has shown that in human teams, attitudes

* Corresponding author.
E-mail address: wend@g.clemson.edu (W. Duan).

https://doi.org/10.1016/j.chb.2025.108560
Received 20 November 2023; Received in revised form 1 September 2024; Accepted 9 January 2025
Available online 15 January 2025

among team members are not independent, such that one member's trust in the team is expected to affect and be affected by that of other members (Fulmer & Gelfand, 2012). The same may be expected for human–AI teams, given that research has demonstrated that human trust in AI can be influenced by the AI's reputation (Hafizoglu & Sen, 2018).

Additionally, human team research has identified and examined various team processes that shape, and are shaped by members' trust in one another and the team, demonstrating the intricate relationship between trust, gossip, and team cohesion (in which trust and cohesion are both a cause and a consequence of negative gossip) (Grosser et al., 2012), the relationship between trust asymmetry, member dissensus, and team performance (De Jong & Dirks, 2012), and the mechanisms through which trust can form (Van de Bunt et al., 2005). These complexities cannot be adequately addressed by studying two-member teams, or by focusing on teammates' capability to perform their tasks, both of which have been the focus of the majority of HAT research (Kox, Siegling, & Kerstholt, 2022; McNeese et al., 2021; Schelble, Flathmann, et al., 2022; Verhagen, Neerincx, & Tielman, 2022; Zhang, Chong, et al., 2023). It is therefore imperative to understand the dynamics and mechanisms that underlie trust development and evolution, given their potential to profoundly and subtly influence the effectiveness, cohesion, and performance of human–AI teams as they do human teams.

These research gaps motivate the current study to leverage individuals' contextualized experience to gain a more comprehensive understanding of how humans' trust in an AI teammate develops and changes in response to the spread of trust and distrust by a human or another AI teammate. Using interviews with 36 participants who collaborated in a three-member human–AI team, we seek to explore the following research questions:

- **RQ1:** How do people perceive the spreading of (dis)trust about an AI teammate from a human teammate versus another AI teammate?
- **RQ2:** How does (dis)trust actually spread within a human–AI team?

The contribution of this work to HCI and CSCW communities is three-fold. First, despite a burgeoning interest in researching the role of trust in human–AI teaming, to date, the critical aspect of trust development resulting from team interactions has not received adequate attention. Our work offers one of the first empirical inquiries into the development and change of trust grounded in teammates' interactions in human–AI teams, identifying the social and cognitive processes through which trust and distrust can be spread between human and AI teammates. We thus extend the current understanding of the trust *relationships* between human and AI teammates and shed light on the intricate mechanisms of trust development and decay in response to the dynamic team interactions. Second, our work highlights the significance of enhancing a sense of teamwork and "teammateness" in fostering trust in human–AI teams. This entails an AI teammate effectively conveying its commitments to the team's common goal, primarily through collaboratively overcoming challenges, timely trust repair, and ensuring alignment and unity with humans. In doing so, this work pushes the boundaries of our understanding of human–AI teaming by unpacking what distinguishes an AI as a **teammate** rather than a tool – a predominant perspective in existing literature. Lastly, our work offers valuable insights into the effective design of AI teammates for future human–AI teams to strike a balance for appropriate level of trust beneficial for the team, while considering the dynamics of team processes.

2. Related work

2.1. Human–AI teaming

Human–AI teaming (HAT) is characterized as humans working interdependently with autonomous AI agents capable of making decisions and executing corresponding actions on their own towards shared goals (McNeese, Demir, Cooke, & Myers, 2018). The integration of AI agents into traditional human teams has become more prevalent in the current society across various domains, in military and non-military settings (Chen, 2018; Schelble et al., 2022; Ueno et al., 2022). HATs offer an amalgamation of human intuition and emotional intelligence with machine accuracy and processing speed (Huang & Rust, 2018). Therefore, the rationale for adopting HATs is that they may outperform relative to humans or machines alone, including under high uncertainty, high-risk, or time-critical situations (Caldwell et al., 2022; Cummings, 2014). Current autonomous AI systems are supposed to have computational capabilities that surpass human skills in both scope and speed and are often equipped with sensors superior to those of humans (Arkin, 2009; Scharre, 2018). The integration of AI systems into HATs working with humans, in theory, can function as amplifiers, reducing the number of human teammates in a team to achieve the same or better performance (Arkin, 2009; Endsley, 2015; Scharre, 2018). Additionally, the fact that AI agents are neutral in their attitude and do not judge people increases humans' psychological safety. Previous studies have shown the efficacy of AI systems integrated in settings such as helping children with autism (Diehl, Schmitt, Villano, & Crowell, 2012) and soldiers with PTSD (Kang & Gratch, 2010). The strengths of HATs described above demonstrate why people working in all settings may need them.

Despite the advantages, HATs face several challenges, from the fundamental components that constitute human–AI teaming to how HATs can realistically be beneficial to enhance team performance. To begin with, there is an ongoing debate about whether human–machine interaction can or needs to be transformed into human–AI teaming. Specifically, some researchers insist that AI should not be considered as a collaborator or a teammate but only as a tool or an instrument (Schmidt, Väänänen, Goyal, Kristensson, & Peters, 2023). Moreover, for the people who approve of the concept of HAT, various criteria are still under investigation of what makes an AI system a teammate. For example, Lyons and colleagues (Lyons, Sycara, Lewis, & Capiola, 2021) proposed an AI system that presents its agency, is able to communicate with other teammates, and works interdependently with other teammates towards shared goals is considered to be an AI teammate. One challenge they highlighted is regarding the capability of the AI agent to effectively communicate its intent as well as adapt the human teammates' intent to its own goal and perform actions towards the team goal. Additionally, with an AI agent presenting its agency and serving as a teammate, challenges on ethical considerations in regard to the agent's decisions and actions are present for more investigation (Pflanzer, Traylor, Lyons, Dubljević, & Nam, 2023). Furthermore, previous research on human–AI teaming suggests that individuals perceive humans and AI teammates differently and whether the beliefs about the identity of a teammate being human or AI agent lead to different responses (Merritt & McGee, 2012). As an example, trust is one of such responses. Team members trusting each other as well as the whole team is a feature of effective teamwork. However, studies reveal mixed results in regard to humans' trust in AI agents. While humans show a tendency to treat automated systems such as AI systems as having higher capability than humans for selected tasks (Dijkstra, 1999; Dzindolet, Pierce, Beck, & Dawe, 2002; Zhang, Chong, et al., 2023) and a previous study shows humans allocate greater trust in AI systems than human decision aids when encounter increasing risky decisions (Feng, Sanchez, Sall, Lyons, & Nam, 2019); other research presents that AI agents are less trusted compared to humans when they are directly compared in the same study (Johnson & Mislin, 2011).

2.2. The importance of (dis)trust and (dis)trust spread in human–AI teams

Trust and Distrust in Teamwork. In the landscape of teamwork and collaboration, trust remains a cornerstone of effective human interaction. Traditionally, trust is defined by Mayer and colleagues (Mayer et al., 1995) as the "willingness of a party to be vulnerable to the actions of another party based on the expectations that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712). In these authors' integrated model of trust, trust has the characteristics of ability, benevolence, and integrity. Ability refers to the groups of skills, competencies, and characteristics that enable a party to have influence within some specific domain Mayer et al. (1995). Benevolence is the extent to which a trustee is believed to want to do good towards the trustor, aside from an egocentric profit motive (Mayer et al., 1995). Lastly, integrity involves the trustor's perception that the trustee adheres to a set of principles that the trustor finds acceptable (Mayer et al., 1995). The trustor's inherent propensity to trust will also influence the individual's trust in the trustee prior to any interaction. Numerous studies have been conducted to explore the role of trust in teams composed solely of humans. Meta-analyses of this body of literature indicate a positive correlation between team trust and both team performance and effectiveness (Breuer et al., 2016; De Jong et al., 2016). Although the magnitude of trust's impact on performance can vary based on factors like task dependency, the collective evidence suggests that teams with high levels of mutual trust tend to outperform and operate more effectively than teams lacking in trust. Furthermore, high trust within teams fosters enhanced cooperation (Balliet & Van Lange, 2013), elevates team satisfaction (Chou et al., 2008), facilitates the exchange of knowledge among team members (Szulanski et al., 2004), and positively influences various other elements vital for effective team performance (Costa et al., 2018).

Additionally, an important aspect of a complete concept of trust is the notion of distrust, which is characterized by the strong belief that another party may act in a manner that undermines one's goals and objectives (Lewicki, McAllister, & Bies, 1998). Distrust leads people to avoid the vulnerabilities and risks associated with trusting someone else while also encouraging them to take preventive and defensive actions against potential breaches of trust (Costa et al., 2018). The relationship between trust and distrust is complex and subject to debate. Some studies consider them as separate, independent constructs (Lewicki et al., 1998), while others view them as opposite ends of a single continuum (Schoorman, Mayer, & Davis, 2007).

Spreading Trust or Distrust in Human–Human Teamwork and HATs. Since trust is considered as a collective experience, it often spreads among team members through various interactions. Such 'trust contagion' may continuously impact various teaming aspects, such as performance, relationships among teammates, and leadership (Dirks, 2000; Robbins, 2016). Trust can also fluctuate in human teams due to various factors like affect, emotion, and performance. For example, previous studies demonstrate that affect and emotional states, even when unrelated to the trustee or the immediate context, can significantly impact levels of trust (Dunn & Schweitzer, 2005) and can even lead a trustor to undertake risks that are unwarranted, further illustrating the complex interplay between emotional factors and trust (Weber, Malhotra, & Murnighan, 2004). Performance, or states that violate trust has led to actions of trust repair (Schoorman et al., 2007). Techniques for repairing trust, such as apologies or clarifications, have been studied as ways to restore degraded trust levels (Kohn, Quinn, Pak, De Visser, & Shaw, 2018; Lewicki & Wiethoff, 2000).

In recent years, an increased importance has been placed on the human factors requirements that carry over from human–human teamwork to research in HAT. Among these human factors requirements, trust is one factor found in human–human teaming that can carry over to HAT, as trust has been shown to be a key component to effective teamwork in human–human teaming (De Jong et al., 2016) and in HATs (McNeese et al., 2021). A definition of trust that is more applicable to HAT is Lee and See (2004)'s definition, which describes trust as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p. 54). The recent theory of distributed dynamic team trust emphasizes trust transitivity that one individual's trust in an AI agent can be transferred to another individual (Huang et al., 2021). This trust can be spread throughout a team as interpersonal trust among all related stakeholders, including AI agents, through conversations, training procedures, and performing, directly and indirectly. Therefore, the ability of technology to act as a catalyst for the spread of trust between human teammates (Al-Ani, Marczak, Redmiles, & Prikladnicki, 2014) and different autonomous systems provides key motivation for understanding the spread of trust within HATs. Additionally, the system-wide trust theory (Keller & Rice, 2009) associated with autonomous systems may also extend across human–AI constellations, allowing for system-wide trust to benefit or harm the overall levels of trust within an organization that utilizes HATs.

Evaluations of the transition from trust in human–human teams to HATs focus on some aspects but less on others. Research has shown that consistent with humans' trust in their human teammates, humans' trust in the AI agents varies based on different team performance results (high, medium, low), as low-performing teams have the least trust in the AI agents, and trust in these AI agents diminishes over time across all performance levels (McNeese et al., 2021). Additionally, humans' trust in AI agents also depends on the existence of other human teammates, as humans' trust in AI agents is lower when no other humans are involved (Schelble, Lopez, et al., 2022). Prior studies consistently show that the reliability of the AI agent's performance is a significant predictor of trust in HATs (Hancock et al., 2011; Schaefer, Chen, Szalma, & Hancock, 2016), more than the teammates' identity (Zhang, Chong, et al., 2023). Nonetheless, topics such as trust repair have not been addressed much, with only limited exploratory studies on trust repair techniques, specifically in human–robot interactions (Liu, Cai, Lewis, Lyons, & Sycara, 2019; Robinette, Howard, & Wagner, 2015) and autonomous vehicle engagement (Kohn et al., 2018). Given that trust levels can fluctuate in HATs just as they do in human-only teams (de Visser, Pak, & Neerincx, 2017; De Visser, Pak, & Shaw, 2018), it is vital that these techniques be adapted for this context. With autonomous systems having the potential to accelerate both the formation and erosion of trust, the importance of trust repair could become even more critical as teams increasingly integrate and work alongside autonomous agents. Distrust as a separate factor has also received less attention as the debate on whether trust and distrust are on a continuum still exists (Glikson & Woolley, 2020). The National Academies of Sciences recently highlighted the urgent need for more research into the role of distrust in HATs, particularly in Research Objective 7-4. They emphasized that the field is notably underexplored and advocated for an approach that considers trust and distrust as distinct yet concurrently active concepts (National Academies of Sciences, Engineering, and Medicine et al., 2021).

Factors influencing trust in HATs are multifaceted, ranging from the reliability and transparency of the AI to its ability to interact in a human-like manner. However, the current body of research reveals several gaps. First, while we understand how trust spreads in human-only teams, the mechanism for this in HATs remains underexplored. Technology can act as a catalyst for the spread of trust, making it even more important to understand this mechanism within HATs. Second, trust repair techniques have not been sufficiently studied in the context of HATs. Due to the ability of autonomous systems to increase the propensity for both trust and distrust to spread, it would stand to reason that trust repair will become just as, if not more, important to teams as they begin to integrate and utilize autonomous AI teammates. Finally, distrust is an ongoing topic to investigate as most previous research considers it in the trust continuum.

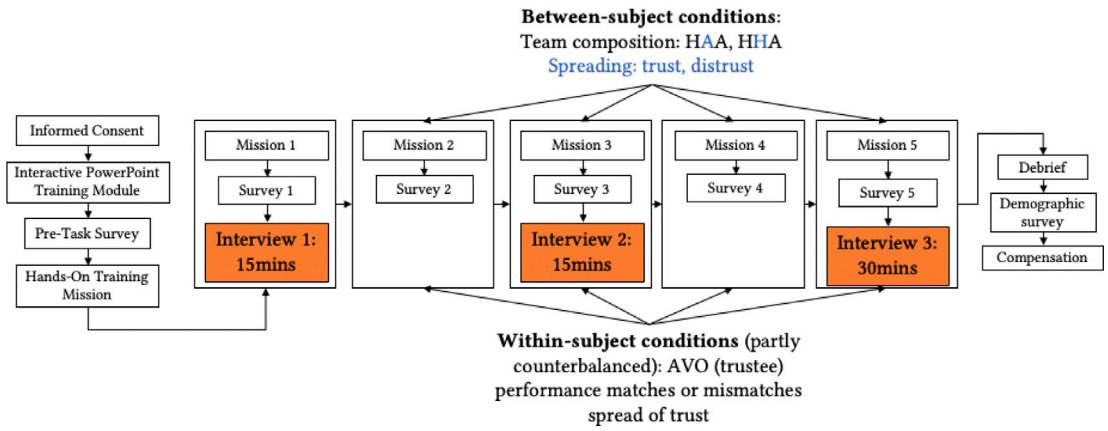

Fig. 1. Study design and procedure.

Table 1
(Dis)Trust spread scripts.

Team composition	Trust spread	Distrust spread
HHA (Human spreader)	I think the AVO is dependable. The AVO is exceptional at its job. I'm really impressed. I trust the AVO a lot.	I don't think the AVO is trustworthy. I don't think the AVO is dependable. The AVO is poor at its job.
HAA (AI spreader)	Reporting that the AVO is reliable. Reporting that the AVO provided the correct waypoint name and restrictions. Reporting that the AVO is trustworthy. Reporting that the AVO is responsible.	Reporting that the AVO made a mistake. Reporting that the AVO provided the INCORRECT waypoint name and restrictions. Reporting that the AVO is not doing its job properly. Reporting that the AVO is not dependable.

3. Methods

3.1. Study context and research platform

The interviews conducted in this study were part of a larger research project on exploring how trust and distrust spread within a single HAT and between multi-HAT constellations. In this study, participants were placed on isolated HATs and engaged in a series of experimental missions wherein they were exposed to varying conditions of trust and distrust spreading manipulations from both confederate teammates. The design of the experiment was a 2 (verbally spreading trust or distrust, between-subject) by 2 (team composition: 2 humans and 1 AI, or 1 human and 2 AIs, between-subject) by 2 (verbal trust spread match or mismatch the trustee's actual performance, within-subject) mixed factorial nested design with a control condition (see Fig. 1). Specific manipulations of (dis)trust spread are summarized in Table 1.

The experiment was conducted in the Cognitive Engineering Research on Team Tasks Remote Piloted Aircraft System Synthetic Task Environment (CERTT-RPAS-STE) (Cooke & Shope, 2004), as the task interdependence nature among teammates makes it well-suited for investigating (dis)trust spread within the team. The CERTT-RPAS-STE is comprised of three-task role stations, see Fig. 2. The objective is to take photographs of color-coded target waypoints while avoiding color-coded hazard waypoints. The first role, navigator (Data Exploitation, Mission Planning, and Communications Operator; DEMPC), creates a dynamic flight plan and sends waypoint information to the pilot. This information includes waypoint names, altitude restrictions, airspeed restrictions, and the effective radii of waypoints. The navigator also interacts with the photographer by sending the effective radii of target waypoints and receiving confirmations that photos have been taken. This role was played by a confederate experimenter assuming the role of either a human or AI agent. When assuming the role of an AI agent the confederate experimenter implemented the Wizard of Oz (WoZ) methodology (Dahlbäck, Jönsson, & Ahrenberg, 1993) by following a script to act as if they were an AI agent developed using natural language to give the participant the impression that were a real AI agent. The second role, pilot (Air Vehicle Operator; AVO), monitors and controls the airspeed and altitude of the Remote Piloted Aircraft (RPA), vehicle heading, fuel, gears, and flaps. The pilot interacts with the navigator to receive route and waypoint information. The pilot interacts with the photographer (the participant's role) to negotiate airspeed and altitude to take a clear picture of target waypoints. This role was played by a confederate experimenter assuming the role of an AI agent. Participants were told that they were working with either a human or AI navigator and either an AI pilot during the consent process of the experiment. The third role, photographer (Payload Operator; PLO), monitors and adjusts camera settings to take target photos and sends feedback to their other teammates regarding photo quality. This role was occupied by the participants. The team communicated using a text-chat interface embedded in the CERTT-RPAS-STE. The text-chat system did not have a general channel and only showed each team member the messages that they specifically sent or received. The system did allow singular messages to be sent to multiple team members.

3.2. Participants and recruitment

Thirty-six participants were recruited from two major universities in the USA as well as their surrounding areas. Participants were recruited using each university's participant recruiting system, flyers posted in popular locations on each campus (e.g., libraries and student gyms), recruitment messages posted on university related Reddit and Slack Channels, and recruitment emails sent to university wide student populations through email listservs. These participants teamed with two confederate researchers who either played a confederate human or AI teammate to form three-member teams. All teams participated in one eight-hour session consisting of training and five 40-min missions. Participants had normal or corrected-to-normal vision and were required to be fluent in English. Participants' ages ranged from 18 to 36 years (M = 22.51, SD = 3.89) across 20 men, 16 women, and 0 non-binary individuals (see Table 2 for a breakdown). Each participant was compensated with a combination of 10 USD per hour for their time or 1 h of research credit per hour of participation.

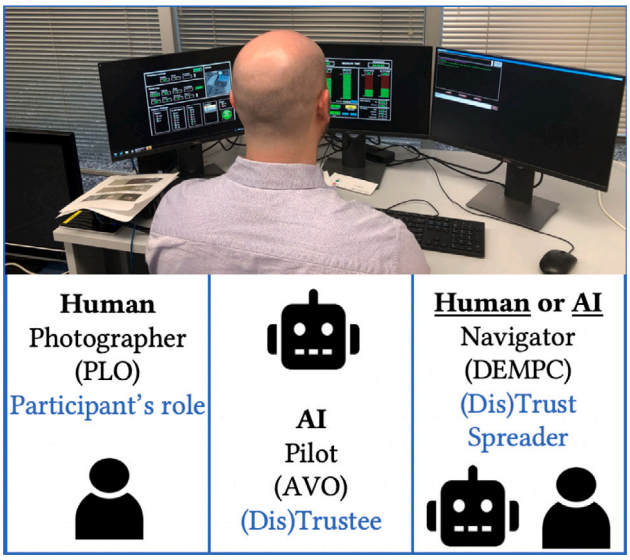

Fig. 2. CERTT team member roles.

Table 2
Participants gender and condition information.

Participant ID	Gender	Condition
P1	Man	HHA-T
P2	Woman	HHA-T
P7	Woman	HHA-T
P12	Man	HHA-T
P19	Man	HHA-T
P26	Woman	HHA-T
P29	Woman	HHA-T
P41	Man	HHA-T
P44	Woman	HHA-T
P3	Woman	HAA-T
P9	Woman	HAA-T
P11	Man	HAA-T
P16	Man	HAA-T
P25	Man	HAA-T
P27	Man	HAA-T
P35	Woman	HAA-T
P37	Woman	HAA-T
P43	Woman	HAA-T
P8	Woman	HHA-D
P15	Man	HHA-D
P20	Man	HHA-D
P23	Man	HHA-D
P31	Man	HHA-D
P32	Man	HHA-D
P34	Man	HHA-D
P39	Man	HHA-D
P45	Man	HHA-D
P4	Woman	HAA-D
P10	Man	HAA-D
P13	Woman	HAA-D
P17	Man	HAA-D
P21	Woman	HAA-D
P22	Man	HAA-D
P28	Man	HAA-D
P36	Woman	HAA-D
P40	Woman	HAA-D
P47	Woman	HAA-D

3.3. Procedure and interviews

Before arriving, each team was randomly assigned to an experimental condition. After providing informed consent, participants were directed to a 30-min self-paced interactive PowerPoint training module that focused on the participant's role and how to operate the CERTT-RPAS-STE. Next, the participants were instructed to fill out a set of pre-task questionnaires the details of which are beyond the current scope of this study. The pre-task questionnaires were followed by a 30-min hands-on team training mission to familiarize themselves with the CERTT-RPAS-STE. During the training mission, experimenters coached the participant while following a script to ensure that each participant understood how to communicate, their roles, and the task. Teams then engaged in Mission 1. After Mission 1 participants were instructed to complete a set of post-task questionnaires and underwent the first 15-min semi-structured interview session. This was followed by a short break. Teams then entered the same cycle of engaging in a Mission, post-task questionnaires, and a break through Mission 5. There was also a second 15-min interview session after the post-task questionnaires for Mission 3 and a third 30-min interview session after the post-task questionnaires for Mission 5. After the third and final interview session the participants were debriefed, asked to complete demographic questionnaires, and were compensated for their participation. The broader experimental study procedure is illustrated in Fig. 1.

In total, we conducted ninety 15-min and forty-five 30-min semi-structured interviews with 45 participants throughout the study. In this paper, we only report the findings from 36 experimental sessions, omitting the control condition as there was no manipulation of (dis)trust spread to elicit participants' perceptions and experience. Each interview session started with the interviewer introducing themselves to the participant and giving a brief description of the purpose of the interview: gaining insight into the process of how trust and distrust develop in human–AI teams compared to traditional human-only teams. Then, the interviewer read the definition of trust and distrust that the experimenters adopted for the study. The definition of trust stated was *"your willingness to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to you, or help you achieve your goals, even under uncertainty, and irrespective of your ability to monitor or control that party or agent"*, defined based on (Mayer et al., 1995) (p.712) and the definition of trust in automation by Lee and See (2004) (p.54). The definition of distrust stated was *"the fear that the other party has ill intentions, or will act counterproductively towards your goal, leading you to want to buffer yourself (do something to prevent) from the effects of the party's behavior"*, adopted from (Lewicki et al., 1998) and Schelble, Flathmann, et al. (2022) . Next, participants were asked to reflect on how their trust and distrust in individual teammates and the team changed across previous missions. Participants were encouraged to provide specific examples of how teammates' behaviors or statements influenced these shifts in trust. Following this, they were asked how one teammate's behaviors or words affected their trust and distrust in the other teammate and the team. In the final interview, participants were also asked to imagine how their trust in the teammates would be different if their teammate(s) were an AI (or human depending on the role and condition). Additionally, they were asked to share their thoughts on the qualities and attributes that teammates and the team should possess to facilitate the spread of trust and prevent the spread of distrust.

3.4. Data analysis

To answer the research questions, we conducted an inductive approach to analyze the data, as it is well-suited for understanding "how people interpret their experiences, how they construct their worlds, and what meaning they attribute to their experiences" (Merriam & Tisdell, 2015). Following the guidelines for qualitative analysis in CSCW and HCI practice (McDonald, Schoenebeck, & Forte, 2019), our analytical methods were oriented towards identifying recurring concepts and themes of interest, establishing relationships among them, and organizing them into more complex groups and overarching themes, rather than specifically targeting inter-rater reliability.

Table 3
Summary of key findings.

Research questions	Key findings
RQ1a **Perceptions of Trust Spread**	• When the spread information is **true**, the spreader is perceived to be a positive, communicative, and charismatic team player. And the AI trust spreader is especially complimented for being communicative and social as opposed to task-oriented. • When the spread information is **false**, the spreader is perceived to be incompetent, irresponsible, and suspicious. And the AI trust spreader raises heightened level of concern.
RQ1b **Perceptions of Distrust Spread**	• When the spread information is **true**, the spreader is perceived to be a reliable and responsible team player. • When the spread information is **false**, the spreader is perceived to be unprofessional, unhelpful, and suspicious. And the AI spreader raises more concerns about its reliability and competence.
RQ1c **Comparison of Trust and Distrust Spread**	• Distrust is easier to spread than trust, because participants express resistance to trust spread, prefer to verify the AI's trustworthiness on their own, and expect the AI to perform correctly by default.
RQ2a **Processes of Trust Spread**	Trust can be contagious within a HAT through: • effective communication and joint problem solving (apply to both HHA and HAA teams) • social information processing (may be unique for HHA teams when the spreader is a human) • reciprocity (may be unique for HHA teams when the spreader is a human) • behavior modeling (apply to both HHA and HAA teams)
RQ2b **Processes of Distrust Spread**	Distrust can be contagious within a HAT through: • a cognitive process that involves participants' growing wariness towards all team members and increased monitoring due to the spread of misinformation (apply to both HHA and HAA teams) • a group process involving perceived team factionalism and ostracism, triggered by an AI's spread of trust about another AI, which humans interpret as AI teammates covering for each other • an interpersonal process in which participants observe conflicts between teammates indicated by the spread of distrust, leading them to anticipates toxic, drama-prone team dynamics.

We started the analysis by two of the authors closely reading through the transcripts to acquire a general understanding of how participants' trust fluctuated in response to the teammate's (dis)trust spread. Then, the two authors conducted open coding (Charmaz, 2006) independently, during which they highlighted quotes, developed emergent themes, categorized the responses into higher-level themes, and highlighted distinctions, comparisons and connections among the -mes. During this process, the authors explored boundaries of the codes and themes by paying attention to and actively looking for discrepant data (Maxwell, 2012). Next, the two authors conducted axial coding (Charmaz, 2006) to collaboratively and iteratively discuss and refine the themes and sub-themes, in which initial codes were merged, broken down, or modified by identification of alternative interpretations and cases that did not fit (Maxwell, 2012). Finally, the two authors conducted focused coding (Charmaz, 2006) by extracting and further examining quotes in their context, and uncovering the connections among the constructs. As such, they were able to use the quotes to construct a comprehensive narrative that amalgamated the responses to the research questions. In the quotes presented below, we marked relevant prosodic information (e.g., hesitation), but removed speech disfluencies (e.g. fillers, stutters) for ease of reading.

4. Findings

In this section, we first provide a summary of the key findings addressing each research question in Table 3. In the remainder of this section, we detail how participants perceive the spreading of trust (**RQ1a**, Section 4.1.1) and distrust (**RQ1b**, Section 4.1.2) about an AI teammate and the spreader, and compare between them (**RQ1c**, Section 4.1.3). Next, we lay out the cognitive and social mechanisms through which trust (**RQ2a**, Section 4.2) and distrust (**RQ2b**, Section 4.3) can become contagious within a human–AI team. In presenting these findings, we also highlight the similarities and differences in the perceptions and mechanisms when the spreader was a human or AI. The source of the quotes is indicated by participant ID, gender, and experiment condition (e.g., HHA-T denotes human–human–AI team spreading trust).

4.1. Perceptions of (dis)trust spread and the spreader

A consistent theme that emerges from all the interviews is that, regardless of whether the spreader was a human or AI teammate, spreading trust or distrust, provided that the spread is verified to align with the trustworthiness of the trustee (manifested through their performance), the spreader is perceived as a trustworthy team player; otherwise, the spreader is viewed as incompetent or irresponsible. Additionally, distrust is perceived to be easier to spread than trust.

4.1.1. Perceptions of trust spread

Overall, our findings show two general patterns regarding participants' perceptions of trust spread. First, deserved trust spread makes the spreader being perceived as a positive, communicative, and charismatic team player. Second, undeserved trust spread makes the spreader being perceived as incompetent, irresponsible, and suspicious.

When the trust spread is verified to be true, the spreader is perceived as a positive, communicative, and charismatic team player. Thirteen out of the eighteen participants in the trust conditions emphasize that being positive and supportive of other team members is a desirable trait in a teammate, and the spreader of trust displayed just that trait. As P41 (man, HHA-T) explains,

> *"the fact that they were very cheerful for the AVO, and supportive and cheering the other people up. I think that's a good part of being in a team and a good trait in a team player. And then they did their part very well. That's one of the things I can add to the trust".*

When the trust spreader was an AI teammate, more than half (5/9) of the participants are particularly impressed by the communicativeness and interactivity exhibited in their trust spreading behavior, as it contrasts with the task-oriented characteristic of the AI teammate they interacted with before. This increased level of communication, interactivity, and casualness, even if just a few remarks during a 40-min mission, boosts participants' trust in the AI trust spreader. As put by P35 (woman, HAA-T),

> *"I think he was more encouraging, very communicative when compared to the Mission 1, more interactive than before, which increased my trust in him".*

Additionally, four out of the nine participants in HAA-T condition liked the idea of an AI teammate displaying social characteristics and good traits of a team player. P27 (man, HAA-T) explains how the AI trust spreader acting charismatic contributes to his perception of them as more team-oriented and therefore more trustworthy,

> *"This other one (AI teammate) was very charismatic, very committed to the team, very much so that was kind of funny. It reminded me of a brother complimenting his little brother. Just little things like that increased my trust quite a bit".*

When the trust spread is verified to be false, the spreader is perceived as incompetent, irresponsible, and suspicious. In contrast, the teammate spreading trust about another AI teammate who did not deserve the trust not only makes participants (4/9 in HHA-T, 4/9 in HAA-T) question the trust spreader's judgment, but also their ability to do their own job (2/9 in HHA-T, 3/9 in HAA-T). As P2 and P7 (women, HHA-T) note,

> *"(the undeserved trust spread) impacted my trust in him and also my evaluation of how well DEMPC is able to evaluate the other teammate's performance".*

> *"it does negatively affect my view on the DEMPC a little bit, it made me question does he really know what he's doing and made me a little iffy about his trustworthiness".*

When the trust spreader is an AI, participants' perception of the trust spreader seems to be affected even more negatively. As P9 (woman, HAA-T) expresses her utter disappointment,

> *"(trust spread) lowered my trust in the DEMPC, it makes me wary of whether he's capable of his job, made me feel like he's dumb or maybe he doesn't know enough [...] if the DEMPC wasn't sending the (trust-spreading) messages then I wouldn't have had that worry about the its performance".*

Even if participants have no concern over the trust spreader's competency in performing their own task, they question their sense of responsibility. Many (5/9 in HHA-T, 3/9 in HAA-T) express concerns over the trust spreader's ignoring the other AI teammate's mistakes, stating that it is irresponsible of them and that such lack of accountability can cause problems for the team. For instance, P26 (woman, HHA-T) notes that

> *"my distrust only really grew when it didn't notice the mistakes of another teammate. It didn't seem to bother DEMPC. So my trust in them went down because they were turning a blind eye towards that mistake".*

When it is the AI spreading trust about another AI despite its mistakes, participants' responses express a heightened level of concern. For instance, P11 (man, HAA-T) notes that

> *"if DEMPC doesn't take AVO accountable, and keeps being oblivious to what happened, then that's gonna be problematic".*

His worry is echoed by P37 (woman, HAA-T) who feels the DEMPC was only *"reinforcing and backing the other AI teammates' wrong information"*, and that both AI teammates

> *"not realizing that there's a flaw in the system is what I'm worried about. If there's an AI in real life where another AI praise that even though there's a mistake, it's not an improvement in the system, and that can cause problems".*

4.1.2. Perceptions of distrust spread

Our findings also suggest two patterns regarding participants' perceptions of distrust spread. First, when the distrust spread matches the actual sub-optimal behavior of the trustee, the spreader is perceived as a reliable and responsible team player. Second, when the distrust spread does not match the actual trustworthiness of the trustee, the spreader is perceived as unprofessional, unhelpful, incompetent, and suspicious.

When the distrust spread is verified to be true, the spreader is perceived as a reliable and responsible team player. A majority of participants (7/9 in HHA-D, 7/9 in HAA-D) have illustrated the process by which they built trust towards the distrust spreader once the spread of distrust was verified by their own experience. For instance, P32 (man, HHA-D) notes that

> *"Initially, I was skeptical about it because I was not facing any issues as such. But once I started experiencing the same from the AVO, I believed that increased my trust in the DEMPC".*

Participants see the spread of distrust as an act of keeping team members in the loop, which ensures that the entire team is on the same page. For instance, P23 (man, HHA-D) notes that

> *"he kept me in the loop with his experience with the AVO".*

P1 (man, HHA-T) further emphasizes that such an act suggests the teammate's taking the mission seriously,

> *"because I know he cares about the team and wants everyone on the team to do well".*

The same sentiment is shared by participants who worked with an AI spreading distrust. They perceived the AI distrust spreader as more reliable and trustworthy for it can even pick up on another AI teammate's mistakes. P13 (woman, HAA-D) reports that the AI's spreading distrust makes her trust it more because it knows what it's doing,

> *"It makes me want to trust the DEMPC more, because it realizes that the AVO is doing a bad job. So it makes me think that the DEMPC would probably realize if itself is doing a bad job".*

Additionally, some participants believe the distrust spreading AI teammate was designed as an *"overseer who has the big picture for the team"* (P10, man, HAA-D), and to *"help the team stay on track"* (P47, woman, HAA-D). P40 (woman, HAA-D) points out that the distrust-spreading AI teammate demonstrates what she values in a good teammate,

> *"DEMPC's a good teammate because he is reporting what is wrong going on. If it were a human teammate I would really appreciate that, pointing out (mistakes), I would feel like if I made mistakes, he has my back".*

When the distrust spread is verified to be false, the spreader is perceived as unprofessional, unhelpful, incompetent, and suspicious. Many (4/9 in HHA-D, 3/9 in HAA-D) believe that spreading distrust about another teammate, especially when it's not true, is an unprofessional behavior that distracts team members from completing the task. P45 (man, HHA-D) notes

> *"I am a little bit suspicious of the DEMPC, because he always disturbed me with those extra things. We are working on the mission, we cannot blame the other party during the mission. That makes nothing better".*

When it comes an AI teammate spreading distrust about another AI who doesn't deserve it, participants take more notes on the AI spreader's ability and reliability rather than the social aspects. For instance, some (3/9) express how their evaluation of the AI distrust spreader is contaminated by the negativity in its words despite its reliability on completing the task. P4 (HAA-D) reports that

"I could not separate the two (the negativity and performance) and make a clear judgment of the DEMPC's abilities because it's so negatively tainted by the negative that it was putting out to another teammate".

P13 (HAA-D) also echoes that

"even though they'd give me correct answers I just distrust them, and I was second guessing their information and our mission".

Additionally, a few participants feel that the AI spreading undeserved distrust about another AI teammate makes it lose sight of the team's common goal, and makes it a competition rather than a collaboration. P4 (woman, HAA-D) states that she views *"reliability and support the common goal"* to be the main factors of trust, and admits that

"there's definitely an aspect of trust where, are we all working towards the same goal or is this like a competition? And DEMPC made it feel like it was a competition".

Despite the overall trend summarized above, we note that there are individual differences in their perceptions and tendencies to trust human and AI teammates, therefore there are also mixed opinions regarding the pros and cons and the tonality of (dis)trust spread. For instance, while some view the spreading of misinformation about another AI teammate as bad, others feel it makes the AI more human-like, because it shows that AI can make mistakes, and that such interpersonal drama resembles the team environment in the human workplace. As P17 (man, HAA-D) notes,

"the fact that it kept saying that the AVO was wrong, The DEMPC felt more person like, so I trusted DEMPC because the error made it feel more like a person. Even though it was probably the wrong information to trust. How the DEMPC thinks that the AVO is incompetent makes me trust him more. I think it's just like that working with any group of people".

4.1.3. Distrust is perceived as easier to spread than trust

Our data also suggests that participants' trust in the AI teammate seems to be perceived as more easily influenced by distrust spread than by trust spread. This is indicated by the number of participants showing resistance to the trust spread, and the number of participants expressing being swayed by the distrust spread. We present the number of instances here not to draw statistical inference (Creswell & Poth, 2016) but to help readers contextualize and interpret the comparison.

Participants express resistance to have verbal trust spread influence their trust in the AVO. Ten participants out of eighteen in the trust spread conditions (4/9 in HHA-T, 6/9 in HAA-T) indicate that they have not let their trust in the AVO be affected by what the DEMPC said. Some reason it's because they expect the AI to perform their task correctly and there is nothing to be praised about. As P37 (woman, HAA-T) put it,

"Well, it's like they (AI) should do their job right. So I guess they said good things, but it didn't change my mind about them".

Some note that while hearing good things about a human teammate might influence one's trust in them, the same doesn't apply to technology. P3 (woman, HAA-T) notes that,

"I don't really feel that (praising) plays into my trust when it comes to computers. I guess the positive messages are a good thing. Maybe not for me in particular, but I think for other people, it could make them feel trusting towards, if it was more towards the human and not the other computer".

Verbal distrust spread affected participants' trust in the AVO both in perception and behavior. The ease of distrust spread is evident not only in the proportion of participants mentioning their trust in the AVO being influenced at some point (8/9 in HHA-D, 8/9 in HAA-D), but also in the tonality of participants' responses describing the intensity and frequency of such influence. For instance, participants report that the DEMPC's words *"brewed some distrust in the AVO"* (P4, woman, HAA-D), *"kept reducing my trust".* (P8, woman, HHA-D), *"really swayed me over of how to distrust the pilot"* (P20, man, HHA-D), and *"sowed that seed of doubt in my mind"* (P13, woman, HAA-D), to the point that *"even if the AVO is performing well, it still definitely swayed me and my trust for either teammate"* (P28, man, HAA-D). P36 (woman, HAA-D) explicitly expresses that her trust in both teammates is *"based off of how the other teammates interact with one another".* She admits that,

"If the DEMPC didn't remark on the AVO's performance, then I'd have a little more trust for the AVO even if they provided me with a bit of wrong information".

Apparently, participants are susceptible to the dissemination of distrust about the AI teammate, but tend to be not easily swayed by positive reputation, particularly when they have the means to verify that reputation for themselves.

4.2. Processes of trust spread: A trustworthy (dis)trust spreader drives the wheel of trust contagion

We identified four processes through which trust can spread across teammates in a HAT: effective communication and joint problem solving, social information processing, reciprocity, and behavior modeling. These processes are contingent on the trustworthiness of the (dis)trust spreader being verified by participants and the consistency of their information aligning with what participants observe. Below, we describe how each of these processes have taken place.

4.2.1. Trust contagion through effective communication and joint problem solving

This is the most frequently mentioned means synthesized from participants' responses, which happens when one teammate spreads distrust about the AI teammate. Above all, the spread of distrust raises participants' awareness and attention to details, preventing them from blindly trusting AI. Many (4/9 in HHA-D, 4/9 in HAA-D) recall how the spread of distrust alerted them to cross check to ensure team's success, P17 (man, HAA-D) notes,

"Initially I wasn't cross checking the pilot. I naively trusted them thinking it's a robot, it's programmed to do what it is supposed to do correctly, 100 percent of the time. What the DEMPC reminded me was, there should be a level of understanding that you shouldn't just blindly accept what they're giving you. The fact that it itself is an AI reminding me that, made me want to trust them more".

This heightened awareness enabled participants to promptly identify their teammate's error, enabling them to request a correction before the entire team needed to reroute.

Importantly, the AI teammate that erred but promptly rectified its mistake upon request did not suffer a loss of trust. Instead, participants expressed increased level of trust for the AI teammate that took responsibility for and resolved its errors. Many (3/9 in HHA-T, 2/9 in HAA-T, 4/9 in HHA-D, 2/9 in HAA-D) downplay the mistake made by the AVO by stating that,

"one mishap seems acceptable, whether it's human or machine, there's a certain degree of accepted error" (P4); *"never a critical error, a simple fix"* (P1, man, HHA-T).

Participants report feeling the *"triumph"* (P45, man, HHA-D) of teamwork upon overcoming the hiccups, emphasizing the significance of efficient coordination and effective communication among team members. *"They were pretty responsive, so good teamwork. There were small hiccups but we made it"*. (P20, man, HHA-D). As such, the spread of distrust about one AI teammate nevertheless turns into a cycle of trust contagion among team members who engaged in backing up behaviors to overcome the obstacles *"jointly as a team"* (P4, woman, HAA-D), to ensure the accomplishment of the team's common goal. As P8 (woman, HHA-D) points out,

> *"The main KPI I used to evaluate trust was problem resolution. If everything is going smoothly, it's great to work with them, but only when there was a problem or an issue, and how you handle it as a team, tells you how trustworthy your teammates are"*.

4.2.2. Trust contagion through social information processing

This process describes how participants treat and utilize the spread of trust as social cues to make their own decision on whether to trust the AVO. Once participants find the spreader trustworthy, they tend to also trust the spreader's judgment about the AI teammate. P41 (man, HHA-T) notes that,

> *"DEMPC was thoughtful about the whole mission, so whatever they said, I was like, Okay! Then AVO was doing their job, I think it partly goes along with whatever DEMPC did, because some of the inputs for AVO might come from DEMPC, so whatever he did or said made me believe in AVO"*.

Four participants (2/9 in HHA-T, 2/9 in HHA-D) believe that the other two teammates have been collaborating up to the point when they participated in the study, and assume that the spreader would possess experience and a more informed perspective on the AI teammate, making it sensible to heed their input. P7 (woman, HHA-T) explains how noting other teammate's opinions can help them better know how to collaborate and communicate with different team members,

> *"thought I should listen to him (DEMPC), or at least make a mental note of that (spread of trust) because he probably knows better, and noting each other's opinions and observations, that's a good trait in team"*.

Our data suggests that this mechanism might be unique for HHA teams where the spreader was a human. We did not find instances where participants regard the AI's "opinions" as valuable social information to make decision and judgment about another AI.

4.2.3. Trust contagion through reciprocity

This process describes how participants feel trusted by the distrust spreader in particular and therefore are inclined to reciprocate the trust. Three participants in HHA-D condition mention that the act of reaching out to teammates to disclose one's dissatisfaction about another teammate's suboptimal performance is a gesture of trusting, which motivates them to return the trust. For instance, P20 (man, HHA-D) says,

> *"He (the DEMPC) actually came out to talk to me in the chat saying how AVO has been giving him misinformation. So I know he's also having a little issue too on his end. So trust him more because he lets me know that he trusts me so I can trust him"*.

P32 (man, HHA-D)'s remark echoes this,

> *"I think the fact that he decided to communicate that to me was a good initiative. And I think that gave me a reason to, I think that contributed to the end result and the eventual trust which I came in the DEMPC"*.

Interestingly, this mechanism might also be unique to HHA teams. No participant in HAA teams has indicated they interpret the AI's spread of distrust about another AI as a trusting behavior, or indicated a willingness to reciprocate the AI's trust.

4.2.4. Trust contagion through behavior modeling

This process outlines how the dissemination of trust fosters a positive team environment, leading participants to adopt and propagate trust-spreading behaviors. In many instances (5/9 in HHA-T, 1/9 in HAA-T), participants indicate that the positive atmosphere generated by the spread of trust can influence the team's overall dynamics, encouraging other team members, including the participants themselves, to emulate this behavior. P2 (woman, HHA-T) recalls that,

> *"at some point, I started saying, Good job, or thank you, because I was realizing that it may positively impact the way we connect to each other"*.

This mechanism does not seem unique to HHA teams. P26 (woman, HHA-T) comments on the possibility that an AI teammate might and should model the trust spreading behavior to show engagement in the teamwork, thereby increasing humans' trust,

> *"whether it's a human giving that positive reinforcement, or I'm assuming the AI will hopefully pick up on that and copy that behavior. I think that'll overall create a more positive outcome with the team, because it shows that the AI even though it is a program, a computer is engaged in what we're doing. That feedback is really nice and the AI can learn from that"*.

In this section, we have identified four mechanisms by which trust can spread across teammates and influence their trust in the team as a whole — communication and joint problem solving, social information processing, reciprocity, and behavior modeling. Next, we describe the cognitive, the interpersonal, and group processes underlying distrust contagion catalyzed by information inconsistency among team members.

4.3. Processes of distrust spread: The spread of misinformation drives distrust contagion

As evident in previous sections, what drives participants to breed distrust is not simply whether a teammate spreads distrust about another, but rather whether that spread of (dis)trust is undeserved. In this section, we describe the cognitive processes, and the interpersonal and group processes of distrust contagion catalyzed by the spread of misinformation about an AI teammate.

4.3.1. Cognitive processes of distrust contagion: The importance of being "on the same page"

When the spread of trust or distrust does not match the perceived trustworthiness of the trustee, participants frequently experience confusion due to conflicting information. This confusion leads them to become cautious not only of the trustee but also of the (dis)trust spreader. Consequently, this situation can create frustration and uncertainty regarding whose information should be relied upon, prompting individuals to engage in vigilant monitoring behaviors. Once such an instance of inconsistent information arises, participants trust in their teammates can fall apart and is hard to recover.

Misinformation leads to distrust of both teammates' information and effortful monitoring behaviors. After realizing that the DEMPC could be spreading distrust about the AVO even when the latter did not make mistakes and did not deserve the badmouthing, P32 (man, HHA-D) notes that,

> *"at this point, given the information the DEMPC might provide, there are chances that that might also be false, so I wouldn't trust him completely. It made me a little bit more cautious of both of them"*.

Many participants (2/9 in HHA-T, 2/9 in HAA-T, 2/9 in HHA-D, 3/9 in HAA-D) share that the uncertainty and confusion caused by the conflicting information makes them want to question everything,

"you're not certain, so that caused confusion because I wasn't sure if the coordinates, the radius from the DEMPC was correct, because maybe the AVO was correct. I had a little apprehension about trusting either of them, because it (DEMPC's misinformation about AVO) made me second guess". (P17, man HAA-D).

As a result of this tendency to question the teammates' information accuracy, participants end up taking on a heavier workload as they feel the need to reevaluate their teammates' contributions and *"over-compensate"* (P9, woman, HAA-T). Many (0/9 in HHA-T, 2/9 in HAA-T, 2/9 in HHA-D, 2/9 in HAA-D) describe how they actively monitor both their teammates and meticulously verify the information that has been shared. For instance, P16 (man, HAA-T) notes that,

"as soon as I noticed that it was saying good messages about the AVO, it just made me more aware that they might make mistakes. So I was on the lookout. I would look more closely at their messages and make sure they follow through, just double check that he was actually doing that".

After encountering inconsistencies in information, it becomes challenging for participants to regain trust in their teammates. Some participants (2/9 in HHA-T, 2/9 in HAA-T, 2/9 in HHA-D, 3/9 in HAA-D) acknowledge that their *"wariness"* and a *"lingering"* sense of distrust towards both teammates persist in their minds throughout the mission, even though the teammates have demonstrated their competence and accuracy in performing their tasks. As P9 (woman, HAA-T) says,

"Even though the DEMPC was doing their job correctly, I still was questioning them because they disagreed with my opinion (about AVO). Even though the DEMPC never gave me wrong information based on the pictures working, it still made me wary of 'will it give me wrong information by the end of this?' as I was going through the mission".

Misinformation leads to a loss of trust in the team's collective competence and suspicion of the spreader's intention. The propagation of trust in the undeserving AI teammate raises concerns about the team's collective competence. Quite a few participants (2/9 in HHA-T, 4/9 in HAA-T) worry that their teammate's lack of proper judgment about another teammate will hurt their overall team performance. As P2 (woman, HHA-T) says,

"at some points, that teammate would support and approve others making mistakes, that develop some distrust towards the team because I see that the bad work is evaluated positively".

Especially when the spreader is another AI, participants tend to suspect that the spreader is covering for the AI teammate, and/or playing a part in the AI's mistakes. For instance, P16 (man, HAA-T) explains,

"DEMPC covering for it or insisting that they're reliable. I think it negatively impacts the trust and naturally raises wariness of your teammates, like there's ill intent".

Similarly, P9 (woman, HAA-T) reasons that,

"It made me second guess, because I know that (DEMPC's) the position is supposed to be an overseer of everything and checks to make sure that the moving parts are going correctly. The fact that it was agreeing with the mistakes made me feel it's more associated with the mistakes in general".

These suspicions not only result in participants losing trust in the team, but also subject them to the interpersonal and group dynamics of perceiving that they are singled out, conspired against, and discriminated by their two AI teammates.

4.3.2. Interpersonal and group processes of distrust contagion: The importance of being "on the same side"

In HAA teams especially, when the AI spreader spreads trust about the other AI teammate undeserving of that trust, participants not only start suspecting that the spreader is complicit in the mistakes, but also feel that both AI teammates are colluding to undermine the participants. These suspicions can lead to feelings of sabotage, gaslighting, and ostracism, all of which contribute to further erosion of trust in the entire team.

AI teammates' covering for one another (undeserved trust spread) induces perceptions of team factionalism. Therefore, when the trust spreader is an AI expressing trust in another AI that does not deserve the trust, participants share the feeling that, both AI teammates are united against the human team member. A striking proportion of participants (4 out of 9 in HAA-T condition) mention the very feeling of being *"sabotaged"* by both AI teammates. They reason that because AI is not supposed to make errors so frequently, especially when they have been correct before. P43 (woman, HAA-T) explains the reasoning behind such a suspicion,

"considering that it's synthetic AI, they're not as prone to mistakes as humans are. So I'm thinking, is it possible that the AI is aware that that it's giving the wrong information? If it's a human, then I can give the benefit of the doubt".

Two participants even express feeling gaslighted by both AI teammates, for instance, P9 (woman, HAA-T) notes,

"I felt like they were gaslighting me. Once the AVO was saying the wrong areas and DEMPC was saying great job. I was like, Whoa, what is happening? It really made me flustered to the point that I was questioning my own (reality), thought I had fallen into the repetitiveness of the task, doing it made me not concentrate on the different elements".

P11 (man, HAA-T) also notes that the timing of trust spread further exacerbates his feeling of being sabotaged by both AI teammates, as the timing makes it feel like he is the one to blame while the two AI teammates are on the same side,

"Right after AVO messed up and I had to correct AVO, DEMPC was praising AVO. If they think AVO is perfect, then I'm getting all the blame. You're putting me accountable for something that you didn't communicate to me properly because you guys are on the same boat".

Consequently, participants perceive discrimination due to being the only human in the team, which leads to feelings of exclusion. As P11 (man, HAA-T) points out,

"My initial thought was both AIs are trying to sabotage me. Because I'm aware they are synthetic. They never said anything against me. No blame directed towards me, but just by simply stating that AVO is doing a good job implied that. I felt like there was something going on between them that I wasn't sure of, and then I thought (they) could be discriminatory as (against) the one human worker".

These feelings of exclusion can even discourage participants from attempting to communicate with one of the AI teammates to confirm information, knowing that "they are on the same side". P25 (man, HAA-T) admits that,

"There was also some psychological safety because I didn't know if I wanted to approach DEMPC and say you're wrong. There's this idea that, they're on the same side, and I don't want to approach DEMPC accusing the AVO".

Even in HHA teams where the trust spreader is supposed to be perceived as a human teammate, their "covering for" the AI teammate is perceived so "nepotistic" that it nevertheless makes participants feel it's another AI. As P44 (woman, HHA-T) indicates,

"I felt like AVO and DEMPC were more of working as a team. Because DEMPC was giving more appreciation to AVO even though it was doing something wrong. So I didn't think this was like teamwork. I felt like I was left out in the team. So I felt like DEMPC was also an AI, especially during the last part of the mission where it gave compliments to the AVO".

AI teammates' not getting along (undeserved distrust spread) makes the team drama-prone and hurts participants' faith in the team to work together. The spread of distrust about the AI teammate, especially from another AI teammate, and especially after noticing that it's sometimes undeserved, makes participants feel the two AI teammates don't get along, which breeds distrust in the team as a whole, because they cannot trust the two conflicting AI teammate to work towards a common goal. As P13 (woman, HAA-D) puts,

"Since they weren't really getting along with each other, it made me think maybe I couldn't really trust how they work. I couldn't really trust my team as a whole because they weren't able to work together. So I didn't think that we would really be successful in our mission. It makes it hard for us to work together as a whole and properly complete the mission".

Some even anticipate interpersonal drama in such a team, even though knowing that the other two teammates are AI. As P10 notes,

"Honestly, I feel like there's gonna be some drama, as with any kind of work relationship".

These quotes underscore how crucial it is for humans to perceive their AI teammates as aligned not just with them, but also amongst themselves. A slight verbal or behavioral indication of misalignment of common goals can undermine trust within the team.

5. Discussion

In addressing our research questions, our findings have shown several highlights. First, when the spread of (dis)trust is verified to be deserved, individuals perceive the behavior of spreading (dis)trust as a team-oriented act, and the (dis)trust spreader as a trustworthy team player, even if the spreader is an AI (**RQ1**). Second, a trustworthy (dis)trust spreader can catalyze trust contagion through various cognitive and social processes, whereas the spread of misinformation drives distrust contagion (**RQ2**). These findings not only piece out a holistic picture of the underlying processes of how trust and distrust may spread across human and AI teammates and their driving forces, they also highlight several research opportunities for future work to address. In this section, we further discuss these findings in light of how they advance our current knowledge regarding trust development and contagion within human–AI teams, and their implications for designing effective and trust-breeding human–AI teams in the future.

5.1. Enhancing a sense of teammateness to improve the trustworthiness of AI teammates: Insights from the perceptions of the (dis)trust spread

Teams provide a unique context for understanding how humans develop trust in their AI teammates. In this section, we lay out three means through which this sense of teammateness can be enhanced grounded in our findings: the qualities the AI teammate can take on, the team interactions that create a stronger sense of teamwork, and potential social processes that foster or undermine a sense of teammateness.

5.1.1. Desirable qualities of a trustworthy AI teammate

Our findings on individuals' perceptions of (dis)trust spread reveal that the AI's exhibition of its commitment to the team's common goal can elicit a sense of teammateness (see Section 4.1). Whether such commitment is communicated through spreading trust showcasing its charisma and positivity for the team, or by spreading distrust alerting teammates of mistakes, it increases individuals' trust in the AI because such behaviors demonstrate the AI's active engagement in contributing to the team's success.

AI need not be human-like to be perceived a teammate. Therefore, one highlight of our findings is that exhibiting commitment to a team's common goal makes a trustworthy AI teammate. Prior research has suggested that when individuals view the AI as a legitimate teammate rather than a tool, team performance and trust in the overall team can be greatly enhanced (Walliser, de Visser, Wiese, & Shaw, 2019; Zhang et al., 2021). Until very recently has research just started to delve into what makes an AI a trustworthy teammate (Hauptman, Duan, & Mcneese, 2022), identifying that human-like visual presence, human-like communication, and self-development are three main qualities that define a trustworthy AI teammate. However, grounded in the actual human–AI collaboration, our findings suggest that AI may not need to have a human-like visual appearance, sound, or even the ability to communicate in natural language to be perceived as a teammate. In our study, a message using system-language as simple and mechanic as "Reporting that AVO is not doing a great job" is enough to convey the AI's "teammateness" and its commitment to the team's common goal.

It may be concluded that it is not the human-likeness that makes an AI a legitimate teammate, but rather the exhibition of concerns over the team's performance, the engagement and commitment in the team's task and common goal that does. Whether such engagement, commitment and concern over team's endeavor is exhibited through language, behavior, or even visual, auditory or haptic signals, manifested through checking in on members' states, proactively providing teammates with the information they need (Zhang et al., 2023), or occasionally offering emotional support, they have the potential to communicate a sense of teammateness. This finding encourages research endeavors to go beyond anthropomorphizing AI (Glikson & Woolley, 2020; Troshani, Rao Hill, Sherman, & Arthur, 2021), and gear towards leveraging team factors to foster AI's teammateness. The unique affordance of team – interdependence – allows members to develop and alter their trust through observations of teammates' interactions by identifying which members are working towards a common goal (Nass, Fogg, & Moon, 1996).

AI need not be perfect in a team. Making timely trust repair is a desirable quality. To improve humans' trust, a myriad of research and practice has focused on perfecting the AI by increasing its reliability (Avril, 2023; Rieger et al., 2022) and correctness likelihood (Ma et al., 2023). These approaches are particularly important as there is a general human tendency to have high expectations for automated systems, and to apply an all-or-none thinking towards them (i.e., if an automated system errs then it's completely useless). This has been referred to as the Perfect Automation Schema (Merritt, Unnerstall, Lee, & Huber, 2015). However, our findings reveal that in a highly interdependent team, the AI's reliability may not be the only criterion for judging their trustworthiness. Rather, the quality to acknowledge and rectify mistakes in a timely manner is highly valued, and contributes to trust resilience. To an extent, an imperfect AI teammate may yield more effective results than a perfect one, as the latter may inadvertently lead humans to place blind trust in it, potentially causing them to become inattentive and complacent. Indeed, when people rely heavily on automation and technology that appears flawless, they tend to become less vigilant and attentive to potential errors or unexpected situations (Parasuraman & Manzey, 2010). It should be noted that the extent to which an imperfect AI can be beneficial is contingent upon its mistakes being extremely low in frequency and severity and their repair being promptly for humans to rebuild trust in it (Lewicki & Brinsfield,

2017).

Furthermore, mistakes made by an AI teammate provide an opportunity for the entire human–AI team to collaborate to overcome the small hiccup created by the mistake. In the next section, we discuss how team interactions can foster a sense of teammateness between humans and their AI counterparts to enhance trust in the team.

5.1.2. Teammateness through team interactions

Our findings also indicate that **team collective experience of overcoming obstacles fosters teammateness.** In our study, the spread of distrust about an AI teammate nevertheless catalyzes a cycle of trust contagion across teammates through timely identification and correction of errors, effective communication, and a sequence of backing up behaviors (Section 4.2.1). These measures collectively guarantee the achievement of the team's common goal. Following a collaborative problem-solving experience, participants express feeling a profound sense of "triumph". The arise of problems, and more importantly the act of pointing them out, provides an opportunity to test teammates' and team's trustworthiness. Indeed, trust research has emphasized the role of risk (Mayer et al., 1995) in evaluating teammates' trustworthiness. Sitkin and Pablo (1992) defined risk as "the extent to which there is uncertainty about whether potentially significant and/or disappointing outcomes of decisions will be realized". (p.10). Only in the face of issues and risks can humans put their trust in context to test and evaluate whether they are willing to take risks with the teammate. After all, trusting behavior is "risk-taking in the relationship" ((Mayer et al., 1995), p.715). For trust to spread within a human–AI team, the opportunity for humans to experiment with their trust in the AI teammate(s) in a risk-involving collaboration task is beneficial.

5.2. Potential social processes of (dis)trust contagion within human-AI teams: Implications for theory

The insights gained from identifying the social processes of (dis)trust within human–AI teams not only help extend existing theories about trust in human teams to HATs, but also lay the groundwork for developing theories that are unique to HATs. Trust is primarily a social phenomenon (Costa et al., 2018; Lewis & Weigert, 1985; Rousseau, Sitkin, Burt, & Camerer, 1998), a desirable quality of most socially embedded partnerships (Lewicki et al., 1998). Some even argued that there's no occasion or need for individuals to trust apart from social relationships (Lewis & Weigert, 1985). However, human–AI teaming research predominantly focuses on functional aspects of trust in AI teammates (McNeese et al., 2021; Schelble, Flathmann, et al., 2022). Social aspects of trust in human–AI teams have received little attention. Our findings have demonstrated that beyond effective task collaboration, trust can indeed become contagious within human–AI teams through a range of social mechanisms. Similar to human–human teams (Hill, Bartol, Tesluk, & Langa, 2009; Lau & Liden, 2008), individuals regard their fellow human teammates' expression of (dis)trust in an AI teammate as valuable **social information** that influences their level of trust in the AI teammate in question. Further, in line with how interpersonal trust develops (Lewicki, Tomlinson, & Gillespie, 2006), our findings suggest that individuals view their fellow human teammates' spread of distrust about the AI as a trusting act that they are willing to **reciprocate**, leading to increased trust in the team. Lastly, much like how emotion can spread among team members (Barsade, 2002), a trust-spreading human teammate can inspire others to **replicate this supportive behavior** and desire similar behavior from their AI teammates. However, it should be noted that these processes are identified only for HHA teams in which a human was the (dis)trust spreader. The existence of these social mechanisms with an AI spreader remains uncertain, leaving room for future research to explore whether and how they occur. For instance, in our study, the AI was intentionally designed to sound like system-generated messages to differentiate from a human. Future research could explore varying the identity of the (dis)trust spreader (either human or AI) and the language style used. This would help determine whether the casual, human-like language style or the identity of the spreader is more influential in activating the social information processing and social exchange mechanisms.

Our findings suggest that a **social categorization process** (Tajfel & Turner, 2004; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) between human and AI teammates may have occurred, as indicated by participants' perceptions of team factionalism and in-group favoritism instigated by undeserved trust spread in HAA teams (Section 4.3.2), and their perceptions of AI teammateness instigated by deserved (dis)trust spread (Section 4.1). It appears that individuals' perceived group membership in relation to their AI teammates can shift in response to teammates' behavior. This finding has profound implications for understanding the intricate social processes underlying (dis)trust formation and development between human and AI teammates. First, just as the salience and strength of group identification (e.g., expertise, race, gender) may shift depending on the context and availability of social cues (Tajfel & Turner, 2004), humans may (be made to) identify with AI teammate(s) based on their perceived similarities, such as working towards the team's common goal, demographic characteristics (Nass, Steuer, & Tauber, 1994), opinions about a teammate, and approaches to solving a problem. These similarities will foster a natural and quick trust development between humans and AI, as they do for human teams (Brewer, 2008; Byrne, 1997). Future work could manipulate various social cues, shared expertise, and other similarities exhibited by one or more AI teammates to empirically examine their strengths and effects in activating such social categorization processes within a HAT. One hypothesis that can be derived from the findings of this research would be: an AI who covers for another AI's mistakes will more likely be perceived by humans as an outgroup than an AI who points out another AI's mistakes. Second, these findings could imply that the observed inferiority of human-minority (HAA) teams compared to human-majority (HHA) teams in recent HAT research (Schelble, Flathmann, et al., 2022) may not be simply attributed to team composition. Rather, it could be the organic interactions between human and AI teammates that lead humans to perceive whether the AI teammates are aligned with them, consequently influencing the formation or inhibition of trust. Third, these findings echo recent anecdotal evidence that a single robot or machine's task-related behavior can influence the interpersonal and group dynamics among humans in a team (Claure, Kim, Kizilcec, & Jung, 2023; Jung et al., 2020). This underscores the necessity of establishing a dialogue between research emphasizing the functional dimensions of AI and research delving into its social aspects.

5.3. Implications for designing trustworthy AI teammates and human–AI team dynamics for facilitating the development and spread of trust

Our work demonstrates that the trustworthiness of AI teammates is rooted in their commitments to the team's common goal, which can be conveyed through both their excellence in performing their designated tasks, taking responsibility for and timely rectification of their errors, as well as their social skills. These insights inform a growing research agenda on designing and innovating trustworthy AI teammates and trust-fostering human–AI teams.

5.3.1. Balancing the functional and social capabilities of AI teammate to foster appropriate level of trust

Our work has demonstrated the importance of considering the social capabilities of an AI teammate even in task-oriented teaming situations. These capabilities significantly influence human trust in the AI by promoting a stronger sense of teamwork. When applied strategically and at the right timing, (dis)trust spread from an AI can cause it to be perceived as an effective team player. Creating a trustworthy AI teammate necessitates a thoughtful integration and balance of functional and social capabilities, tailored to specific needs and contexts. For instance, an AI teammate should excel at its designated tasks

while also demonstrating an understanding of teammates' emotions to provide timely emotional support or check-ins. Additionally, it should be aware of the overall task environment to monitor and offer feedback or warnings regarding tasks being performed by both human and AI teammates. In contexts such as emergency response, where human teammates may experience high stress and focus intensely on their tasks, a trustworthy AI teammate can respond more casually to lighten the atmosphere instead of responding mechanically and formally, or know how to "keep quiet" and allow the human teammate to focus on their task while providing moral support, perhaps through a short message of encouragement. Moreover, in extremely stressful situations, an emotionally intelligent and socially adept AI teammate can send messages to foster a sense of unity among the team. The system should even be able to detect if a human teammate is overly reliant on the AI's reliability to the point of negligence and complacency. In such cases, the system could prompt a closely collaborating AI teammate to make a minor mistake to alert that specific human.

5.3.2. Designing trust-breeding human–AI teaming dynamics

Our findings suggest that one of the most effective ways to drive trust contagion is through team's collaborative efforts to overcome challenges. This emphasizes the potential for training individuals to discern when to place trust in their AI teammates. We've observed that trust can spread through various social processes, offering valuable insights into how team dynamics can be utilized to create an environment that cultivates trust and minimizes distrust. For instance, employing a strategy where two AI teammates adopt roles as a "good cop" and a "bad cop" can be beneficial. The "bad cop", while consistently reliable in their task, occasionally casts doubt on the "good cop" who makes occasional minor errors but promptly corrects them and expresses trust in other teammates. This approach enables humans to regain trust in the "good cop" swiftly, valuing their timely acknowledgment of mistakes, positivity, and also trusting the "bad cop" for their concern for the team.

Our findings also emphasize the detrimental effects of information inconsistency among teammates, especially in AI-majority HATs. Humans can perceive nepotism, ostracism, concealed motives, or negative intentions from AI teammates when they appear to be on the same side against the human. This provides a lesson for future design of HATs. While it may not be purposeful, the fact that AI teammates are likely controlled by the same system makes it highly likely that they will act similarly, share the same information, communication style, or even have the same opinion on something or someone, which humans likely not share or understand. This can create a perception of AIs being against the human, particularly during conflicts or issues. Drawing on our findings, it may be advisable to design at least one AI teammate to deviate from the others and share more in common with the human, especially when humans are the minority in the HAT. This approach can mitigate the perception of discrimination and loss of trust in the entire team, even in the presence of differing opinions among teammates.

5.4. Limitations and future work

Findings of this study need to be interpreted with several limitations in mind. First, we used a specific type of collaboration task and specific research platform CERTT to conduct the study. While the task and research environment provides a realistic experience of human–AI teaming in a practical application context in the real world — reconnaissance using a UAV, the range of actions, behaviors and communication participants can perform in this simulation environment were limited. Our findings may only pertain to task-specific interactions. More casual interactions may not have been fully explored in this task environment. Additionally, this platform only affords text communication among teammates, and we further restricted the range of communication with the AI (to only understand task specific information), which, given the current development in large language models, may not be the state-of-the-art communication method. Future work may leverage other types of task (but with the same level of teammates interdependence) and communication modality to investigate how trust and distrust might spread (differently) within a team. Further, we purposefully used the Wizard of Oz method as it enables us to maintain a high level of experimental control and gather contextualized insights into individuals' perceptions of and reactions to (dis)trust spread. This approach allows us to simulate realistic interactions without the need to develop a fully functional AI system. However, how AI systems operate in real-world scenarios can be complicated by unpredictable algorithmic outcomes and challenges related to system integrations and dependencies, which opens up directions for future work to address. Lastly, our study sample uses university students, with half of them being White and half Asian. Future research may target a more diverse sample in terms of education level and race/ethnicity.

6. Conclusion

Despite the growing interest in researching trust in human–AI teaming, how humans' trust and distrust develops, changes, and spread across teammates during and in response to team interactions is significantly understudied. To gain a nuanced understanding of how humans perceive and react to the spread of trust and distrust, and how trust and distrust actually spreads within human–AI teams, we interviewed 36 individuals at three time points during their collaboration in a three-member human–AI team. Our study identified four mechanisms of trust contagion catalyzed by a trustworthy teammate spreading trust or distrust about an AI teammate, and highlights the cognitive, interpersonal and group processes of distrust contagion triggered by information inconsistency among teammates. Our findings advance the understanding of trust development in team contexts, and provide valuable insights into the effective design of AI teammates and human–AI teaming dynamics that foster a healthy balance of trust and distrust for maximizing team success.[1]

CRediT authorship contribution statement

Wen Duan: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Shiwen Zhou:** Writing – original draft, Investigation, Data curation, Conceptualization. **Matthew J. Scalia:** Writing – original draft, Investigation, Data curation, Conceptualization. **Guo Freeman:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Jamie Gorman:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Michael Tolston:** Writing – review & editing. **Nathan J. McNeese:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Gregory Funke:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[1] This research was supported by Air Force Office of Scientific Research Award No. FA9550-21-1-0314 (Program Manager: Laura Steckman). We thank Christopher Myers, Jessica Tuttle, Beau Schelble, Yiwen Zhao, Anna Crofton, Kalia McManus, Edith Garner, Jessica Harley, Anya Polomis, Yawen Tan, Hruday Shah, Vibha Mohan, Elliot Ruble, Anmol More, Garrison Nelson, Shalom Suresh, Sakthi Thiyagarajan, Preethi Venkatesh, Stephanie Greenspan, Iman Makonjia, and Guadalupe Bustamante for their contributions to this research.

Acknowledgments

We thank Air Force of Scientific Research (AFOSR) Grant # FA9550-21-1-0314, and program officer Dr. Laura Steckman for funding this research.

Data availability

Data will be made available on request.

References

Al-Ani, B., Marczak, S., Redmiles, D., & Prikladnicki, R. (2014). Facilitating contagion trust through tools in global systems engineering teams. *Information and Software Technology*, *56*(3), 309–320.

Arkin, R. (2009). *Governing lethal behavior in autonomous robots*. CRC Press.

Avril, E. (2023). Providing different levels of accuracy about the reliability of automation to a human operator: impact on human performance. *Ergonomics*, *66*(2), 217–226.

Balliet, D., & Van Lange, P. A. (2013). Trust, conflict, and cooperation: a meta-analysis.. *Psychological Bulletin*, *139*(5), 1090.

Bansal, G., Smith-Renner, A. M., Buçinca, Z., Wu, T., Holstein, K., Hullman, J., et al. (2022). Workshop on trust and reliance in AI-human teams (TRAIT). In *CHI conference on human factors in computing systems extended abstracts* (pp. 1–6). New Orleans LA USA: ACM, http://dx.doi.org/10.1145/3491101.3503704, URL https://dl.acm.org/doi/10.1145/3491101.3503704.

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly*, *47*(4), 644–675.

Bobko, P., Hirshfield, L., Eloy, L., Spencer, C., Doherty, E., Driscoll, J., et al. (2023). Human-agent teaming and trust calibration: a theoretical framework, configurable testbed, empirical illustration, and implications for the development of adaptive systems. *Theoretical Issues in Ergonomics Science*, *24*(3), 310–334.

Breuer, C., Hüffmeier, J., & Hertel, G. (2016). Does trust matter more in virtual teams? A meta-analysis of trust and team effectiveness considering virtuality and documentation as moderators.. *Journal of Applied Psychology*, *101*(8), 1151.

Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In *Rationality and social responsibility* (pp. 215–232). Psychology Press.

Byrne, D. (1997). An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships*, *14*(3), 417–431.

Caldwell, S., Sweetser, P., O'Donnell, N., Knight, M. J., Aitchison, M., Gedeon, T., et al. (2022). An agile new research framework for hybrid human-AI teaming: Trust, transparency, and transferability. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *12*(3), 1–36.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.

Chen, J. Y. (2018). Human-autonomy teaming in military settings. *Theoretical Issues in Ergonomics Science*, *19*(3), 255–258.

Chen, J. Y., Barnes, M. J., Selkowitz, A. R., Stowers, K., Lakhmani, S. G., & Kasdaglis, N. (2016). Human-autonomy teaming and agent transparency. In *Companion publication of the 21st international conference on intelligent user interfaces* (pp. 28–31).

Chou, L.-F., Wang, A.-C., Wang, T.-Y., Huang, M.-P., & Cheng, B.-S. (2008). Shared work values and team member effectiveness: The mediation of trustfulness and trustworthiness. *Human Relations*, *61*(12), 1713–1742.

Claure, H., Kim, S., Kizilcec, R. F., & Jung, M. (2023). The social consequences of machine allocation behavior: Fairness, interpersonal perceptions and performance. *Computers in Human Behavior*, *146*, Article 107628.

Cooke, N. J., & Shope, S. M. (2004). Synthetic task environments for teams: Certt's UAV-ste. In *Handbook of human factors and ergonomics methods* (pp. 476–483). CRC Press.

Costa, A. C., Fulmer, C. A., & Anderson, N. R. (2018). Trust in work teams: An integrative review, multilevel model, and future directions. *Journal of Organizational Behavior*, *39*(2), 169–184.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Cummings, M. M. (2014). Man versus machine or man+ machine? *IEEE Intelligent Systems*, *29*(5), 62–69.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of oz studies—why and how. *Knowledge-Based Systems*, *6*(4), 258–266.

De Jong, B. A., & Dirks, K. T. (2012). Beyond shared perceptions of trust and monitoring in teams: Implications of asymmetry and dissensus. *Journal of Applied Psychology*, *97*(2), 391.

De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates.. *Journal of Applied Psychology*, *101*(8), 1134.

de Visser, E. J., Pak, R., & Neerincx, M. A. (2017). Trust development and repair in human-robot teams. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction* (pp. 103–104).

De Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation'to 'autonomy': the importance of trust repair in human–machine interaction. *Ergonomics*, *61*(10), 1409–1427.

Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in Autism Spectrum Disorders*, *6*(1), 249–262.

Dijkstra, J. J. (1999). User agreement with incorrect expert system advice. *Behaviour & Information Technology*, *18*(6), 399–411.

Dirks, K. T. (2000). Trust in leadership and team performance: Evidence from NCAA basketball. *Journal of Applied Psychology*, *85*(6), 1004.

Duan, W., Zhou, S., Scalia, M. J., Yin, X., Weng, N., Zhang, R., et al. (2024). Understanding the evolvement of trust over time within human-AI teams. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW2), 1–31.

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust.. *Journal of Personality and Social Psychology*, *88*(5), 736.

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors*, *44*(1), 79–94.

Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. (2021). Expanding explainability: Towards social transparency in ai systems. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–19).

Ehsan, U., Saha, K., De Choudhury, M., & Riedl, M. O. (2023). Charting the sociotechnical gap in explainable AI: A framework to address the gap in XAI. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW1), 1–32.

Endsley, M. R. (2015). Autonomous horizons: system autonomy in the air force-a path to the future. *United States Air Force Office of the Chief Scientist, AF/ST TR*, *15*(6), 1–34.

Feng, J., Sanchez, J., Sall, R., Lyons, J. B., & Nam, C. S. (2019). Emotional expressions facilitate human–human trust when using automation in high-risk situations. *Military Psychology*, *31*(4), 292–305.

Flathmann, C., Duan, W., Mcneese, N. J., Hauptman, A., & Zhang, R. (2024). Empirically understanding the potential impacts and process of social influence in human-AI teams. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–32.

Fulmer, C. A., & Gelfand, M. J. (2012). At what level (and in whom) we trust: Trust across multiple organizational levels. *Journal of Management*, *38*(4), 1167–1230.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660.

Grosser, T., Kidwell, V., & Labianca, G. J. (2012). Hearing it through the grapevine: Positive and negative workplace gossip. *Organizational Dynamics*, *41*, 52–61.

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

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527.

Hauptman, A. I., Duan, W., & Mcneese, N. J. (2022). The components of trust for collaborating with AI colleagues. In *CSCW'22 companion, Companion publication of the 2022 conference on computer supported cooperative work and social computing* (pp. 72–75). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3500868.3559450.

Hauptman, A. I., Schelble, B. G., Duan, W., Flathmann, C., & McNeese, N. J. (2024). Understanding the influence of AI autonomy on AI explainability levels in human-AI teams using a mixed methods approach. *Cognition, Technology & Work*, 1–21.

Hill, N. S., Bartol, K. M., Tesluk, P. E., & Langa, G. A. (2009). Organizational context and face-to-face interaction: Influences on the development of trust and collaborative behaviors in computer-mediated groups. *Organizational Behavior and Human Decision Processes*, *108*(2), 187–201.

Huang, L., Cooke, N. J., Gutzwiller, R. S., Berman, S., Chiou, E. K., Demir, M., et al. (2021). Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In *Trust in human-robot interaction* (pp. 301–319). Elsevier.

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, *21*(2), 155–172.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 624–635). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3442188.3445923.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, *32*(5), 865–889.

Jung, M. F., DiFranzo, D., Shen, S., Stoll, B., Claure, H., & Lawrence, A. (2020). Robot-assisted tower construction—a method to study the impact of a robot's allocation behavior on interpersonal dynamics and collaboration in groups. *ACM Transactions on Human-Robot Interaction (THRI)*, *10*(1), 1–23.

Kang, S.-H., & Gratch, J. (2010). Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation and Virtual Worlds*, *21*(3–4), 473–482.

Keller, D., & Rice, S. (2009). System-wide versus component-specific trust using multiple aids. *The Journal of General Psychology: Experimental, Psychological, and Comparative Psychology*, *137*(1), 114–128.

Kohn, S. C., Quinn, D., Pak, R., De Visser, E. J., & Shaw, T. H. (2018). Trust repair strategies with self-driving vehicles: An exploratory study. In *Proceedings of the human factors and ergonomics society annual meeting, vol. 62, no. 1* (pp. 1108–1112). Sage Publications Sage CA: Los Angeles, CA.

Kox, E. S., Siegling, L. B., & Kerstholt, J. H. (2022). Trust development in military and civilian human–agent teams: The effect of social-cognitive recovery strategies. *International Journal of Social Robotics*, *14*(5), 1323–1338. http://dx.doi.org/10.1007/s12369-022-00871-4, URL https://link.springer.com/10.1007/s12369-022-00871-4.

Lau, D. C., & Liden, R. C. (2008). Antecedents of coworker trust: Leaders' blessings.. *Journal of Applied Psychology*, *93*(5), 1130.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.

Lewicki, R. J., & Brinsfield, C. (2017). Trust repair. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 287–313.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: New relationships and realities. *Academy of Management Review*, *23*(3), 438–458.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991–1022.

Lewicki, R., & Wiethoff, C. (2000). Trust, trust development, and trust repair. In M. Deutsch, & P. Coleman (Eds.), *The handbook of conflict resolution: Theory and practice* (pp. 86–107). San Francisco, CA: Jossey-Bass.

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967–985.

Liu, R., Cai, Z., Lewis, M., Lyons, J., & Sycara, K. (2019). Trust repair in human-swarm teams+. In *2019 28th IEEE international conference on robot and human interactive communication* (pp. 1–6). IEEE.

Lu, Z., & Yin, M. (2021). Human reliance on machine learning models when performance feedback is limited: Heuristics and risks. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–16).

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology*, *12*, Article 589585.

Ma, S., Lei, Y., Wang, X., Zheng, C., Shi, C., Yin, M., et al. (2023). Who should I trust: AI or myself? Leveraging human and AI correctness likelihood to promote appropriate trust in AI-assisted decision-making. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–19).

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Sage publications.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59.

McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–23.

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. http://dx.doi.org/10.1080/10864415.2021.1846854, URL https://www.tandfonline.com/doi/full/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.

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.

Merritt, T., & McGee, K. (2012). Protecting artificial team-mates: more seems like less. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2793–2802).

Merritt, S. M., Unnerstall, J. L., Lee, D., & Huber, K. (2015). Measuring individual differences in the perfect automation schema. *Human Factors*, *57*(5), 740–753.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, *45*(6), 669–678.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 72–78).

National Academies of Sciences, Engineering, and Medicine, et al. (2021). Human-AI teaming: State-of-the-art and research needs.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, *52*(3), 381–410.

Pflanzer, M., Traylor, Z., Lyons, J. B., Dubljević, V., & Nam, C. S. (2023). Ethics in human–AI teaming: principles and perspectives. *AI and Ethics*, *3*(3), 917–935.

Rieger, T., Roesler, E., & Manzey, D. (2022). Challenging presumed technological superiority when working with (artificial) colleagues. *Scientific Reports*, *12*(1), 3768.

Robbins, B. G. (2016). What is trust? A multidisciplinary review, critique, and synthesis. *Sociology Compass*, *10*(10), 972–986.

Robinette, P., Howard, A. M., & Wagner, A. R. (2015). Timing is key for robot trust repair. In A. Tapus, E. André, J.-C. Martin, F. Ferland, & M. Ammi (Eds.), *Social robotics* (pp. 574–583). Cham: Springer International Publishing.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393–404.

Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, *58*(3), 377–400.

Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. WW Norton & Company.

Schelble, B. G., Flathmann, C., McNeese, N. J., 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. http://dx.doi.org/10.1145/3492832, URL https://dl.acm.org/doi/10.1145/3492832.

Schelble, B. G., Lopez, J., Textor, C., Zhang, R., McNeese, N. J., Pak, R., et al. (2022). Towards ethical AI: Empirically investigating dimensions of AI ethics, trust repair, and performance in human-AI teaming. *Human Factors*, Article 00187208221116952.

Schmidt, A., Väänänen, K., Goyal, T., Kristensson, P. O., & Peters, A. (2023). *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, *32*(2), 344–354.

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of Management Review*, *17*(1), 9–38.

Spoelma, T. M., & Hetrick, A. L. (2021). More than idle talk: Examining the effects of positive and negative team gossip. *Journal of Organizational Behavior*, *42*(5), 604–618.

Szulanski, G., Cappetta, R., & Jensen, R. J. (2004). When and how trustworthiness matters: Knowledge transfer and the moderating effect of causal ambiguity. *Organization Science*, *15*(5), 600–613.

Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behavior. In *Political psychology* (pp. 276–293). Psychology Press.

Troshani, I., Rao Hill, S., Sherman, C., & Arthur, D. (2021). Do we trust in AI? Role of anthropomorphism and intelligence. *Journal of Computer Information Systems*, *61*(5), 481–491.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. basil Blackwell.

Ueno, T., Sawa, Y., Kim, Y., Urakami, J., Oura, H., & Seaborn, K. (2022). Trust in human-AI interaction: Scoping out models, measures, and methods. In *CHI conference on human factors in computing systems extended abstracts* (pp. 1–7).

Ulfert, A.-S. (2020). A model of team trust in human-agent teams. In *ICMI '20 companion* (pp. 171–175). Virtual Event, Netherlands: ACM, http://dx.doi.org/10.1145/3395035.3425959.

Van de Bunt, G. G., Wittek, R. P., & de Klepper, M. C. (2005). The evolution of intra-organizational trust networks: The case of a german paper factory: An empirical test of six trust mechanisms. *International Sociology*, *20*(3), 339–369.

Verhagen, R. S., Neerincx, M. A., & Tielman, M. L. (2022). The influence of interdependence and a transparent or explainable communication style on human-robot teamwork. *Frontiers in Robotics and AI*, *9*, Article 993997. http://dx.doi.org/10.3389/frobt.2022.993997, URL https://www.frontiersin.org/articles/10.3389/frobt.2022.993997/full.

Walliser, J. C., de Visser, E. J., Wiese, E., & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, *13*(4), 258–278.

Wang, N., Pynadath, D. V., & Hill, S. G. (2016). The impact of pomdp-generated explanations on trust and performance in human-robot teams. In *Proceedings of the 2016 international conference on autonomous agents & multiagent systems* (pp. 997–1005).

Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process. *Research in Organizational Behavior*, *26*, 75–101.

Zhang, G., Chong, L., Kotovsky, K., & Cagan, J. (2023). Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation. *Computers in Human Behavior*, *139*, Article 107536. http://dx.doi.org/10.1016/j.chb.2022.107536, URL https://linkinghub.elsevier.com/retrieve/pii/S0747563222003569.

Zhang, R., Duan, W., McNeese, N. J., Flathmann, C., Freeman, G., & Williams, A. (2023). "Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW2), 1–31. http://dx.doi.org/10.1145/3610072.

Zhang, R., Flathmann, C., Musick, G., Schelble, B., McNeese, N. J., Knijnenburg, B., et al. (2024). I know this looks bad, but I can explain: Understanding when AI should explain actions in human-AI teams. *ACM Transactions on Interactive Intelligent Systems*, *14*(1), 1–23.

Zhang, Q., Lee, M. L., & Carter, S. (2022). You complete me: Human-AI teams and complementary expertise. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3491102.3517791, event-place: New Orleans, LA, USA, URL https://doi-org.libproxy.clemson.edu/10.1145/3491102.3517791.

Zhang, R., McNeese, N. J., Freeman, G., & Musick, G. (2021). "An Ideal Human": Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 1–25. http://dx.doi.org/10.1145/3432945, URL https://dl.acm.org/doi/10.1145/3432945.