Understanding the Evolvement of Trust Over Time within Human-AI Teams

WEN DUAN, Clemson University, USA
SHIWEN ZHOU, Arizona State University, USA
MATTHEW J. SCALIA, Arizona State University, USA
XIAOYUN YIN, Arizona State University, USA
NAN WENG, Clemson University, USA
RUIHAO ZHANG, Arizona State University, USA
GUO FREEMAN, Clemson University, USA
NATHAN MCNEESE, Clemson University, USA
JAMIE GORMAN, Arizona State University, USA
MICHAEL TOLSTON, Air Force Research Laboratory, USA

The success of human-AI teams (HATs) requires humans to work with AI teammates in trustful ways over a certain time period. However, how trust evolves and changes dynamically in response to human-AI team interactions is generally understudied. This work explores the evolvement of trust in HATs over time by analyzing 45 participants' experiences of trust or distrust in an AI teammate prior to, during, and after collaborating with AI in a three-member HAT. Our findings highlight that humans' expectations of AI's ability, integrity, benevolence, and adaptability influence their initial trust in AI before collaboration. However, this initial trust can be maintained or revised through the development of situational trust during collaboration in response to the AI teammate's communication behaviors. Further, the trust developed through collaboration can impact individuals' subsequent expectations of AI's ability and their collaborations with AI. Our findings also reveal the similarities and differences in the temporal dimensions of trust for AI and human teammates. We contribute to CSCW community by offering one of the first empirical investigations into the dynamic and temporal dimension of trust evolvement in HATs. Our work yields insights into the pathways to expanding the methodological toolkit for investigating the development of trust in HATs, formulating theories of trust for the HAT context. These insights further inform the effective design of AI teammates and provide guidance on the timing, content, and methods for calibrating trust in future human-AI collaboration contexts.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Human-AI teaming, Human-agent teaming, Human-autonomy teaming, Trust fluctuation, Trust development, Trust evolvement, Qualitative method

ACM Reference Format:
Wen Duan, Shiwen Zhou, Matthew J. Scalia, Xiaoyun Yin, Nan Weng, Ruihao Zhang, Guo Freeman, Nathan McNeese, Jamie Gorman, and Michael Tolston. 2024. Understanding the Evolvement of Trust Over Time

Authors' addresses: Wen Duan, Clemson University, Clemson, SC, USA, wend@clemson.edu; Shiwen Zhou, Arizona State University, Tempe, AZ, USA; Matthew J. Scalia, Arizona State University, Tempe, AZ, USA; Xiaoyun Yin, Arizona State University, Tempe, AZ, USA; Nan Weng, Clemson University, Clemson, SC, USA; Ruihao Zhang, Arizona State University, Tempe, AZ, USA; Guo Freeman, Clemson University, Clemson, SC, USA; Nathan McNeese, Clemson University, Clemson, SC, USA; Jamie Gorman, Arizona State University, Tempe, AZ, USA; Michael Tolston, Air Force Research Laboratory, Wright Patterson Air Force Base, OH, USA.

ACM 2573-0142/2024/11-ART521
https://doi.org/10.1145/3687060

within Human-AI Teams. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 521 (November 2024), 31 pages. https://doi.org/10.1145/3687060

1 INTRODUCTION

Trust has been found to be a key predictor of effective human teamwork [21], positively related to team performance [8, 24], fruitful collaboration [4], efficient exchange of knowledge [98], and positive attitudes toward the team [18]. With AI technology permeating every part of work and life, causing humans to increasingly rely on AI to make decisions [9, 104, 105] and generate contents [35, 63], it becomes even more necessary to understand humans' trust toward AI to ensure proper, safe and effective use [60]. In particular, as AI becomes further embedded within human-AI teams (HATs), assuming the role of a teammate rather than a tool [103, 113], the level of trust and collaborative synergy between humans and AI could substantially impact the success and efficacy of human-AI partnership.

However, trust is not static, but *dynamically evolves and changes over time* through repeated direct [27, 86, 89, 109] and indirect [62] interactions within a team. In the nascent field of HAT, studies have primarily examined trust as static snapshots from an input-outcome perspective. While extensive research efforts have been dedicated to examining the impacts of humans' trust on their adoption [47] and effective use of AI [5, 60, 64], as well as the factors that increase humans' trust, such as AI's reliability [2, 87], transparency [7, 15], and explainability [41, 112], a process-based approach is rare. As such, little is known about how trust in HATs develops initially, waxes and wanes in response to team experiences, and how trust developed through these experiences influences subsequent human-AI collaborations. Understanding these temporal variations of trust is critical for CSCW research focused on human-AI collaboration, in that it will enable researchers to identify pivotal moments, incidents, and factors that cause trust fluctuation, and help devise effective strategies to cultivate, maintain, and calibrate trust in changing circumstances of human-AI teaming.

Indeed, at different stages of human team formation and interaction, what influences members' trust in one another may vary. For instance, in early stages of team formation, initial trust can be built by importing expectations on the basis of members' background, professional credentials and affiliations [78]. Such initial trust is known as "swift trust", which provides an early cognitive basis for team members to interact and collaborate in the absence of familiarity and past experiences [115]. As team members progress into actual collaborations, initial trust is subject to verification through actions and accumulated experiences [52]. Sometimes high initial trust does not maintain and even deteriorates due to members not living up to others' prior expectations [49, 52]. Additionally, the patterns of trust trajectory have been found to vary based on team composition, how members got acquainted, how they coordinated on tasks, and can be shaped by culture [17]. In addition to direct experience, trust also fluctuates in response to team dynamics, where members observe how teammates interact with one another and act upon incidents in ways that honor or violate their trust [62]. As such, trust is sensitive to these situational cues thereby forming "situational trust" [67]. Following repeated collaboration, team members tend to develop a general inclination to trust or distrust their teammate(s), a phenomenon referred to as "learned trust".

Compared to trust research within human teams, these temporal aspects and team dynamics of trust remain significantly understudied, if not entirely absent, in CSCW research on human-AI collaboration. This gap obscures a holistic understanding of how humans' trust evolves and changes over the course of, and in response to team interactions. To address this gap, in this study, we conducted interviews at three time points with 45 participants who collaborated on a three-member team of different compositions (i.e., human-AI-AI, human-human-AI, and human-human-human). In doing so, we explore the following research questions that target trust perceptions before, during, and after the teaming experience respectively.

- **RQ1:** How do people's prior expectations influence their **initial trust** in AI versus human teammate(s) **before** human-AI teamwork?
- **RQ2:** How is people's **situational trust** in AI versus human teammates fostered **during** human-AI teamwork?
- **RQ3:** How does people's **learned trust** developed **after** collaborating with their AI versus human teammates (or lack thereof) influence their subsequent expectations of the teammates and collaborations with them?

This study contributes to the HCI and CSCW knowledge on human-AI collaboration in several respects. First, we present one of the initial empirical investigations into the temporal dimension of trust, derived from individuals' collaborative experiences within human-AI teams. Our work identifies the prior expectations that influence the formation of individuals' initial trust, the behaviors that nurture the development of situational trust, and the impacts of learned trust on subsequent human-AI teaming. These insights thus help delineate the life cycle of trust evolvement and erosion in response to the dynamic team interactions that either confirm or challenge initial expectations. In doing so, we broaden the existing understanding of human trust in AI teammates by highlighting the discrete elements of trust apparent at various phases of human-AI collaboration and unpacking their interconnected relationships. Second, our work highlights the importance of investigating trust while considering the temporal dimension, illuminating the pathways to enhance the methodological toolkit and develop theories of trust specific for HAT contexts to effectively address the temporal variations in trust dynamics. Lastly, our work offers valuable insights into the effective design trust calibration techniques for future human-AI teams.

2 RELATED WORK

2.1 Human-AI Teaming

Human-AI teaming (HAT) is characterized by one or more humans working interdependently with one or more autonomous AI agents that are capable of independent decision-making and action towards shared goals [13, 14, 93, 99]. HATs combine human emotional and intuitive qualities with the precision and rapid processing abilities of AI technologies [46]. Recent autonomous AI systems are often recognized for their superior computational skills, surpassing humans in both scope and speed in a wide range of important tasks, and are typically equipped with advanced sensory capabilities [1, 91]. The integration of AI into HATs could amplify the team's capabilities, potentially reducing the number of humans in the team while maintaining or enhancing overall effectiveness [1, 31, 91].

Nevertheless, the implementation of HATs encounters several challenges. One such challenge is the ongoing debate over whether an AI system should be considered a team member [65, 83]. Research has suggested that differences in people's perceptions of the teammate being a human or an AI can lead to varied reactions [77]. In this context, we study trust as one of the key reactions. The level of trust is positively related to effective teamwork. Yet, findings on human trust in AI agents show inconsistent results. In some studies, people perceive AI systems as more capable [26, 30, 110] and tend to trust AI systems more than human advisors in high-risk scenarios [33]. On the other hand, some studies indicate that AI agents are trusted less than humans when both are evaluated in the same study [51]. These inconsistencies could result from the varying time points at which trust was measured, given that trust can fluctuate moment by moment [69]. In the next section, we review prior work on the fluctuation of trust in human teams to provide further insights.

2.2 Trust and Its Fluctuation in Human Teams

2.2.1 The Importance of Trust in Human Teams. Mayer and colleagues [69] proposed a definition of trust between humans as "the willingness of a party to be vulnerable to the actions of another party based on the expectation 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' trust framework there exists three factors of trustworthiness that constitute the characteristics of the trustee and ultimately predict the level of trust placed in them by the trustor. The three factors are ability, benevolence, and integrity. Ability refers to the context dependent skills, competencies, and characteristics of the trustee [69]. Benevolence is the trustor's belief that a trustee wants to good towards them aside from profit motive [69]. Lastly, integrity is the trustor's belief that the trustee adheres to the same principles as them or principles the trustor finds acceptable [69].

Moreover, Mayer and Gavin [70] tested these three factors of trustworthiness and discovered that the level of trust in the trustee is affected by the trustor's perception of the trustee's ability, benevolence and integrity, which further influences team trust and team performance. A lack of trust in teams not only undermines team members' ability as well as behaviors, but also has a negative impact on team performance [70]. On the contrary, high trust contributes to team outcomes such as team satisfaction, team information processing and team performance [8]. Indeed, the association between team trust and team performance grows stronger as the team becomes more diverse and task interdependence becomes higher [21, 24].

Trust undergoes a continuous evolution over time through reciprocal interactions within teams [62], and develops through close team cooperation and vigilant monitoring [21]. Teams with high trust are inclined to share information, place confidence in the reliability of others, exhibit a greater willingness to be influenced by other members, which in turn increases team performance and cooperation [109]. In contrast, teams with low trust are more likely to experience conflicts [96], limiting their information processing capacity. This limitation arises as team members shift their focus away from shared tasks, and constrains team cognition as stress and anxiety escalate among team members [50]. In addition, vigilant monitoring positively reinforces team trust, even though monitoring behaviors initially imply a lack of trust [69]. However, in the absence of trust, monitoring behaviors can act as a means to support others to perform tasks and keep on track in order to accomplish shared goals, which in turn contribute to trust [25, 69].

2.2.2 Temporal Variations of Trust in Human Teams. Mayer and colleagues [69] accounted for the time dimension in their framework. Prior to interaction, a trustor's propensity to trust will influence how much trust they place in the trustee [69]. When the interaction between the trustor and trustee begins, [69] argue that the integrity of the trustee will be the most salient. Then, after continuous interactions with the same trustor and trustee, the perceived benevolence of the trustee becomes more prevalent [69, 94]. Lastly, the outcome of each trust scenario feeds back into the next scenario involving the same trustee [69].

Following Mayer and colleagues [69] time dimensions of trust, Marsh and Dibben [67] propose that trust can be categorized into three layers: dispositional, situational, and learned. Applied to human teams, dispositional trust is the psychological disposition or personality trait of a team member to be trusting toward another member, the team, and the system or not [67]. Situational trust is when a team member's trust in another member or the team adjusts in response to situational cues [67]. In this way, situational trust is heavily context dependent and relies on the behavioral interactions between team members and each team member and their environment. Lastly, learned trust is the team member's general tendency to trust or not to trust another team member or the team as a result of continuous interactions with said team member or the team [67]. Dispositional trust, like trust propensity, will determine the amount of trust a human team member will have in

another team member and the team before interaction. Situational trust entails the fluctuation of a team member's trust during each interaction. Finally, learned trust is the team member's trust in another member or the team based on a reflection of one and/or many verbal and behavioral interactions with the same member or team. In our current work, we adopt the terminologies and definitions of **situational trust** and **learned trust** from this framework to refer to the trust fostered during collaboration, and the trust developed at the conclusion of a series of collaboration episodes, respectively.

Team members often form their initial trust based on members' backgrounds, social positions and professional credentials [3, 78]. Such initial trust is referred to "swift trust", which allows team members with less familiarity and experience to interact and collaborate at the very beginning of teamwork [115]. Teams with high initial trust are more connected and socialized at the very early stage of teamwork and tend to have better team performance [49].In addition, due to the dynamic nature of trust, initial trust changes on the basis of team members' accumulated experience and task performance during teamwork [52, 115]. For instance, initial trust can be easily destroyed if team members fail to live up to others' expectations, whereas it can also be maintained or even increased if the team has better performance throughout the collaboration [52]. The pattern of trust development has also been found to vary based on team composition, levels of familiarization, closeness in coordination and different cultural values [17]. Aside from direct experience, trust also undergoes changes in response to team dynamics, such that members observe how teammates interact with each other and respond to the teammates' behaviors that either uphold or breach their trust [62]. Furthermore, at different stages of team collaboration, the factors contributing to trust have been found to impose varying importance. For example, Jarvenpaa and Leidner [49] have identified communication behaviors that foster trust during various phases of team formation. They discovered that behaviors such as social communication and communication expressing enthusiasm are instrumental in fostering trust during the initial stages of group formation. Conversely, predictable communication and prompt responses are crucial for promoting trust during the later stages of team development.

2.3 Trust in Human-AI Teams

As HATs grow in prominence, research on trust within HATs has also received significant attention [40, 74]. The most well-adopted definition of trust in HAT research is defined by Lee and See [60] as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (p. 54). Integrating AI into teams [46, 75] necessitates a deep understanding of how trust functions in such hybrid environments.

In the current study, we integrate the definition of trust in organizations [69] (p.712) and the definition of trust in automation by [60] (p.54), to develop our own definition of trust as *"The 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"*.

Trust is critical to HATs as it is related to team performance and the formation of shared mental models within HATs. This relationship is evident in the way team members' trust in AI agents directly impacts their collaborative efficiency and the team's collective success (e.g., [34, 73, 92]). Moreover, teams exhibiting higher levels of trust are more likely to develop robust shared mental models vital for effective coordination and communication, ultimately leading to improved team performance [92]. As such, trust within HATs is not merely a byproduct of interaction but a foundational element that shapes the team's operational dynamics and ability to function as a cohesive unit.

A constellation of factors also influences trust in HATs, such as team composition, cognitive and affective dimensions of trust, and transparency. Each plays a unique role in shaping human perceptions and interactions with AI. One significant factor is the composition of HATs, where the ratio of humans to AI agents within a team can significantly influence team dynamics [80, 92]. This aspect affects perceptions of AI agents' reliability and trustworthiness, thereby impacting overall team performance and cooperation. Additionally, the level of AI's machine intelligence plays a critical role in trust development. As Chen and Barnes [14] observed, higher machine intelligence, allowing for more autonomous and complex actions, can enhance trust through perceived competence. However, it can also lead to mistrust if AI's actions become overly complicated or unpredictable.

Trust in AI involves both cognitive and emotional dimensions. Cognitive trust is based on a rational evaluation of AI's reliability and competence [43, 60], while emotional trust stems from affective elements such as feelings of safety and emotional attachment [54, 71]. In terms of emotional trust, the physical presence or tangibility of AI influences trust perceptions. Tangible, interactive AI systems are often perceived as more real, enhancing users' understanding and sense of control [56].

Transparency in AI operations is also crucial for building trust, especially in systems using complex methods like deep learning [43]. Understanding how AI makes decisions fosters trust among users. The nature of tasks assigned to AI further contributes to trust development [95]; AI that is effectively used for tasks aligning with its capabilities is more likely to be trusted [39]. Additionally, immediacy behaviors, such as proactive and responsive actions, help build cognitive trust by creating a sense of closeness and setting high expectations for AI performance [37].

These elements collectively emphasize the complex nature of trust in HATs, underscoring the importance of AI's design, capabilities, transparency, and interaction dynamics in fostering effective human-AI collaboration. Furthermore, researchers have also found that trust in HATs is dynamic and evolves over time, influenced by the levels of autonomy and human control over AI agents [11, 41]. For example, dynamic situational awareness (SA) is pivotal in establishing appropriate trust levels within HATs. Trust in automated systems develops gradually, influenced by various elements, but remains heavily context-dependent [43].

Despite the gradually increased amount of literature identifying the importance of trust in HATs, significant gaps remain in understanding trust dynamics within HATs. Most notably, the concept of distributed dynamic team trust [45] suggests that trust in one AI agent can extend to trust in other team members, underlining the interconnectedness of trust in these settings. However, the mechanisms through which trust spreads and fluctuates in HATs are not well understood. In addition to the already existing quantitative measures, there is a need for more qualitative, in-depth studies that explore how trust is perceived, formed, and maintained from the perspectives of human team members [10]. Such studies would provide valuable insights into the psychological processes underlying trust in HATs, offering a more nuanced understanding of these complex team dynamics. Additionally, the role of distrust, often considered in contrast to trust, requires further exploration, particularly as it pertains to HATs [21, 61]. Addressing these gaps is crucial for the effective integration of AI in teams, ensuring that the advantages of HATs are maximized while mitigating the challenges they present (e.g., [83]).

3 METHODS

We conducted semi-structured interviews at three time points during an eight-hour experiment. The qualitative data utilized to address the research questions in this study were derived from a larger research project on trust and distrust spread within a HAT. In this project, participants experienced a series of 5 experimental missions on a three-member HAT with varying team compositions and

other experimental manipulations. All the study design, procedure, recruitment and compensation methods were approved by the University's Institutional Review Board. While this paper did not focus on these experimental manipulations of trust or distrust spread, the manipulations served to contextualize participants' experiences that were grounded in these critical incidents, which allowed us to gain insights regarding the nuances and variations of trust across time frames.

To facilitate a better understanding of the context, in this section, we explain the research platform and simulation task, the experiment design and manipulations, participants, the study and interview procedure, and data analysis.

3.1 Research Platform and Simulation Task

The experiment was conducted in the Cognitive Engineering Research on Team Tasks Remote Piloted Aircraft System Synthetic Task Environment (CERTT-RPAS-STE) [20]. The CERTT-RPAS-STE is comprised of three-task role stations, see Figure 1. The three roles are: a navigator or DEMPC (short for Data Exploitation, Mission Planning, and Communications); a pilot, or Air Vehicle Operator (AVO); and a photographer, or Payload Operator (PLO).

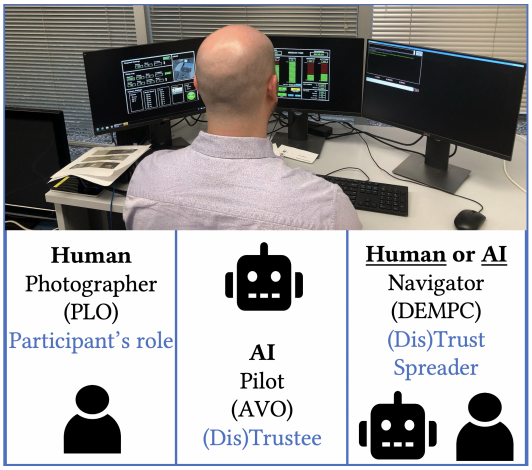

Fig. 1. CERTT Team Member Roles.

The team's task was to take reconnaissance photographs of targets during missions, which required all three team members to communicate using the text chat interface as a feature of CERTT-RPAS-STE. This feature allowed the messages to be sent to one team member privately or to multiple team members simultaneously.

The role of **photographer (PLO)** was always played by a participant, who captured images of the targets and provided feedback on photo quality. The **navigator (DEMPC)** created flight plans and communicated with the pilot about the waypoint information. This role was played by a research confederate using the Wizard of Oz (WoZ) methodology [22], who assumed the role of either a human or AI agent depending on the experiment condition. When assuming the role of an AI, the confederate would strictly follow a script adapted from [88] that documented techniques to create non-human agents in synthetic environments. Additionally, the confederate would respond to the participants only if their message followed the format they were instructed to use, or else it would respond "I don't understand", to simulate an AI that is limited in natural language comprehension. Participants perceived the AI-likeness as we intended them to, as evidenced by their success in

the manipulation check (by choosing the correct teammate role (human or AI)), as well as their interview responses, where they perceived the other teammate(s) as genuine AI. The navigator also communicated with the photographer about effective radii, and received and provided confirmation of the photos. The **pilot (AVO)** controlled the aircraft and communicated with the navigator to receive the waypoint information, and with the photographer to provide target airspeed and altitude, and to receive the confirmation of the photos. The pilot was also played by a confederate using WoZ.

3.2 Study Design and Manipulations

The design of the experimental study was a 2 (Navigator verbally spreading trust or distrust about Pilot, between-subject) by 2 (Team composition: human-human-AI, or human-AI-AI, between-subject) by 2 (Pilot's actual performance: good or bad, within-subject) by 5 (Missions, within-subject) mixed factorial nested design with a control condition (human-human-human with no manipulation). Specific manipulations of (dis)trust spread are summarized in Table 1.

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 trustworhty. 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.

The within-subject manipulation of the pilot (AVO)'s performance was done by adjusting the RPA (remote piloted aircraft)'s altitude, airspeed, and using correct or incorrect waypoint names. For instance, in the "good performance" condition, participants experienced consistent and appropriate altitude adjustments, which allowed them to take clearer photos, whereas in the "bad performance" condition, AVO would make incorrect altitude adjustments and put in wrong waypoint names. Participants would need to identify the mistakes, communicate them back to AVO using the language it could "understand" for it to correct them, to be able to take good photos.

3.3 Participants

Forty-five participants were recruited from two major universities in the USA. Various recruitment strategies were used to diversify the sample, including the universities' participant recruiting system, physical flyers, recruitment messages posted on university Reddit and Slack Channels, and recruitment emails sent through university email listservs. Participants were required to speak and write fluent English. Participants all had normal or corrected-to-normal vision. Their ages ranged from 18 to 36 years (M = 22.51, SD = 3.89), including 25 men, 18 women, and 2 gender non-binary individuals. Twenty-three participants identified as Asian or Asian American, 17 Caucasian or White, 2 identified with more than one ethnic backgrounds or other ethnicity, 1 African American, 1 Hispanic, 1 Native American. Compensation for the participation was offered as either 10 US Dollars per hour or one research credit for every hour of participation. On average, participants

reported to interact with a form of AI on a monthly to weekly basis (M=3.56, SD=1.47) based on their response to the question "How often do you interact with AI (e.g., Siri, Alexa)?" (1=Once a year or less, 5=Everyday, 3=Several times a month, 4=Several times a week).

3.4 Procedure and Interviews

Participants were randomly assigned to one of the experimental conditions. Upon arriving and completion of informed consent, participants were directed to a 30-minute self-paced interactive PowerPoint training that described 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. Then, participants did a 30-minute hands-on team training mission to familiarize themselves with the simulation platform, during which experimenters coached each participant to ensure his/her understanding of how to communicate, their roles, and the task. Teams then engaged in Mission 1 where the pilot and navigator performed no manipulations of trust/distrust spread nor performance match/mismatch. Therefore, participants experienced the same Mission 1 except for team composition. After Mission 1 participants were instructed to complete a set of post-task questionnaires and underwent the first 15-minute semi-structured interview. This was followed by a short break. Teams then went through the same cycle of task collaboration in a Mission, post-task questionnaires, and a break through Mission 5. There was also a second 15-minute interview after the post-task questionnaires for Mission 3, and a third 30-minute interview after the post-task questionnaires for Mission 5. After the final interview, the participants were debriefed about the purpose and manipulations of the study, and informed about our measures of privacy protection, data usage, and their right to have the collected data destroyed. Then they were asked to complete demographic questionnaires, and were compensated for their participation. The broader experimental study procedure is illustrated in Figure 2.

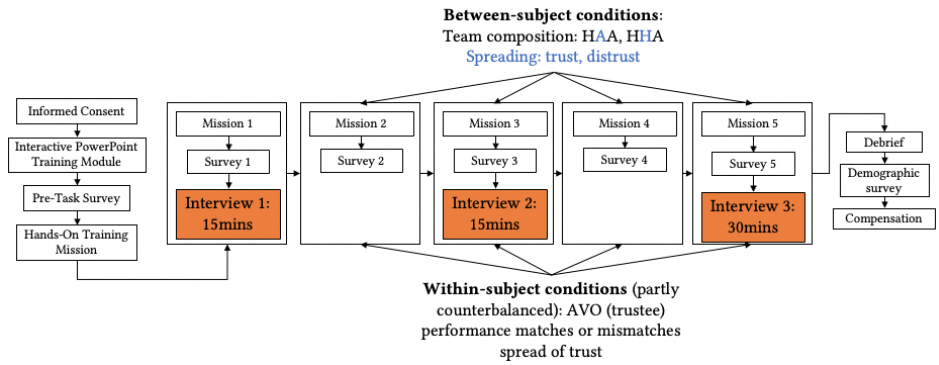

Fig. 2. Study Design and Procedure.

In total, we conducted ninety 15-minute and forty-five 30-minute semi-structured interviews with 45 participants, resulting in over 55 hours of audio recordings. Each interview session started with the interviewer 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-human teams. Then, the interviewer read the definition of trust and distrust adopted for the study. Trust was defined as *"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", by integrating the definition of trust in organizations [69](p.712) and the definition of trust in automation by [60](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 [61] and [92]. Next, participants were asked to reflect on how their trust and distrust in individual teammates changed throughout the mission that just finished or over the course of all previous missions. Participants were encouraged to provide specific examples of how teammates' behaviors or statements influenced these shifts in trust. In all the interviews for the control condition where no manipulation occurred, participants were asked to imagine how their trust in the human teammates would be different if they were AI. In the first interview after Mission 1 for all the other conditions, where the only manipulation was team composition, participants were also asked to imagine how their trust in the human teammate(s) would be different if they were AI, and vice versa, to understand what expectations participants had of AI and humans, prior to interacting with either. In subsequent interviews after Mission 3 and 5, they were asked how their initial trust was maintained or altered through actually collaborating with the teammates, and how their trust might differ if the roles of AI and human were swapped.

3.5 Data Analysis

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" [76]. Following the guidelines for qualitative analysis in CSCW and HCI practice [72], 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.

Two of the authors first closely read through the transcripts to gain a general understanding of how participants formed their trust toward human and AI teammates at different phases of collaboration. Then, the two authors conducted open coding [12] independently, during which they highlighted quotes, developed emergent themes, categorized the responses into higher-level themes, and highlighted distinctions, comparisons and connections among the themes. During this process, the authors explored boundaries of the codes and themes by paying attention to and actively looking for discrepant data [68]. Next, the two authors conducted axial coding [12] 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 [68]. Finally, the two authors conducted focused coding [12] 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.

4 FINDINGS

In this section, we first report the influence of prior expectations of AI on individuals' **initial trust** formed before human-AI teaming (**RQ1**). Then, we identify three ways that **situational trust** is fostered during human-AI teaming (**RQ2**). Lastly, we demonstrate the impact of **learned trust** on subsequent expectations and interactions of human-AI teaming (**RQ3**). In exploring these temporal variations of trust across different stages, we also highlight the similarities or differences of trust dynamics concerning AI teammates compared to those related to human teammates. To facilitate a better understanding of the context, we annotate the source of the quotes by participant ID, gender, experiment condition (e.g., HHA-T denotes human-human-AI team spreading trust), as well as the time point of the interview (e.g., M1 indicates after Mission 1).

4.1 The Influence of Prior Expectations on the Initial Trust Formed toward an AI Teammate before Human-AI Teaming

In this section, we identify four prior expectations that participants reported to have: 1) AI is less likely than humans to make mistakes; 2) AI is less likely than humans to have bias, hidden agenda, or conflicting interest; 3) AI is unable to identify and correct errors embedded in their program; and 4) AI is unadaptable to new and changing situations. We elaborate on how the first two expectations have led to participants' higher initial trust in an AI than a human teammate, and the latter two have led participants to have lower or undetermined about their initial trust.

4.1.1 The expectation of AI to be less error-prone led to higher initial trust in an AI than a human teammate. Prior to the interactions with an AI teammate, a majority of the participants (40/45) indicated that they would trust an AI more than a human teammate, because AI is less likely than humans to make mistakes. For instance, like many participants, P15 (man, HHA-D, M1) reported that *"With the AI maybe more trust than a person because there's no human error, it's an AI."* Some even reported unrealistic expectations that AI should be perfect and better than humans in many ways. As P11 (man, HAA-T, M1) put, *"Apparently they're better than humans at some things, many things. There's this expectation people have of AI that AI should be perfect."* Many others echoed this expectation, expressing a belief that if AI were to be put into teams "to replace a human" (P14, man, control), they have to be free of human error to be able to "add value to the team" (P6, non-binary, control).

Further prompting participants for their reasons for such an expectation revealed that their evaluation of the AI's strengths over humans are related to the context and task environment in which the AI is implemented. After familiarizing with the UAV (unmanned aerial vehicle) renaissance task during training, participants assumed that the AI's accurate and fast computing capabilities made the AI more suitable for the task and therefore more trustworthy than humans. As P13 (woman, HAA-D, M1) put, *"I'd probably trust the AI more (than a human) because from my perspective, they're known for accuracy while humans we make a lot of errors."* P7 (woman, HHA-T, M1) also explicitly delineated the specific boundaries within which AI possessed an edge over humans, *"The ability to compute everything very quickly, which is one thing that it has as an advantage compared to a human. So I would say AI is very well suited for tasks like these."*

Additionally, participants reasoned that AI is less prone to errors and more dependable due to its lack of physiological needs, unlike humans who can can experience distractions and fatigue over time. As P4 (non-binary, HAA-D, M1) noted, *"I think AI has the potential to be more dependable across the board because it doesn't have needs like humans do, where we get tired and need sleep and get distracted.* This advantage is especially needed in an environment in which teammates have to engage in tedious repeated tasks. P19 (man, HHA-T, M1) explained that *"Humans can make mistakes and mess up all the time, specially when we are used to doing what we are doing. But AI had been programmed to do the same thing over and over again and don't get bored."*

4.1.2 The expectation of AI to be less likely to have hidden intention, conflicting interests, and bias led to higher initial trust in an AI than a human teammate. Aside from the expectation that AI is less error-prone than humans, some participants put higher initial trust in the AI because they expected the AI to be unable to possess hidden intention like humans do. P1 (man, HHS-D, M1) explained that he would always have a bit of distrust toward people whom he did not know, whereas for AI, he trusted that its design inherently prevents it from having hidden agendas beyond its programmed functions. *"The AI just does whatever it does, it doesn't really have hidden intentions to it. Technically, the AI should be perfect at whatever it is doing, so you don't need that little bit of distrust that you*

want for humans, as in regards to, they're not gonna have any hidden agenda, because they're just acting along what they're programmed to do."

Some participants also mentioned that humans may have different individual goals and conflicting interests in a team, whereas the AI is supposed to be built to achieve the team's goal. For example, P25 (man, HAA-T, M1) brought up a scenario to illustrate that resource allocation and evaluation mechanisms may cause humans to act against other teammates instead of toward a common goal, which is not the case with AI: *"Humans might be competitive for resources or want to win to get promotion... having that human instead of the second AI could cause complications in what we're trying to accomplish. There could be disagreements in what our actual goal is. And that would be the difference with AI. The goal for the AI would have been set previously to (align with) the team's goal. But humans' minds can change."*

The expectation that AI is less likely than humans to have bias also accounts for participants' higher initial trust in an AI teammate. Several participants mentioned that the quality of AI to be "fair" (P2, woman, HHA-T, M3), "objective" (P46, woman, control, M5) and "impartial" (P38, woman, control, M5) made them perceive the AI to be potentially more trustworthy than humans. P38 (woman, control, M5) related her experience of being treated unfairly by a former teammate of a project team, stating that an AI teammate would not discriminate others by their expertise. *"That person (former teammate) consistently ignored my ideas just because I don't have the same technical background. I would imagine an AI be taking inputs and accepting others' suggestions more equally."* In experimental conditions where the human teammate (DEMPC) spread trust (HHA-T) or distrust (HHA-D) about the AI teammate (AVO), which at times the AVO did not deserve, the word "bias" was frequently brought up. P45 (woman, HHA-D, M5) recalled that *"DEMPC seemed to be biased against the AVO. He keep saying the AVO is incompetent."* She believed that the human DEMPC's act of badmouthing the AI was a manifestation of prejudice. In contrast, when DEMPC was an AI spreading (dis)trust about AVO who did not deserve (in HAA conditions), the AI spreader was perceived as having "misaligned information" (P22, man, HAA-D, M5) instead of bias. After experiencing the DEMPC praising the AVO even when it made mistakes, P2 (woman, HHA-T, M3) responded to the hypothetical question "how would your trust change if the DEMPC were an AI?" by saying *"I guess I will trust him (DEMPC) more because I feel like AI may be more capable of evaluating others."* For P2, the human DEMPC's strong bias significantly impaired his/her impartial judgment of others' abilities, a mistake P2 believed an AI wouldn't make due to its freedom from bias.

4.1.3 The expectation of AI to be unable to identify and correct their errors led to lower initial trust or want to hold off their judgment until they could verify the AI's quality. Several participants reported that they would trust an AI less than a human teammate because they expected that AI could not realize and correct the mistakes embedded in its program, reasoning that if it could, it would not have made the mistake to begin with. P18 (man, control, M5) recalled how he quickly resolved a mistake he made and avoided making more mistakes by frequently double checking, which he assumed an AI would not be capable of, *"Less trust (if human teammates are AI). Those things that human can, like double check and look at the error again. I don't know maybe we don't build those things into into our AIs."* Similarly, P36 (woman, HAA-D, M1) explained, *"A second reason for that (more trust for human than AI) is because I really value the points that people can correct themselves despite the mistakes. But for AI, they probably don't realize (mistakes). Why would they make mistakes if they know there are mistakes?"* This quote represents a common belief among participants that the AI cannot act beyond its programmed parameters to detect errors that should have been avoided in the first place.

The expectation that AI cannot identify and correct its mistakes has also led participants to be undetermined about their initial trust. Some reported that their trust would be contingent upon the absolute quality of the AI. As P2 (woman, HHA-T, M3) noted, *"I would maybe trust him (AVO) less and more (if it were a human). Less because of human error. I might trust him more just because if you're a human, and then things start going wrong, then there's still a hope that the human will realize the error. While with AI it's not going to be the case. All you can count on is the quality of AI, hoping it's always correct."* P2 believed that while humans can self-correct even if they might be more error-prone than AI, AI is fixed and unchangeable. Therefore, it is utterly important for AI to be made error-free. A few expressed that they would be more lenient towards an AI's inability to double-check and correct its mistakes than they would be toward a human. For instance, P7 (woman, HHA-T, M3) reported that *"If it was a human and it didn't double check the mistake, I probably would trust it less than an AI. If it was an AI, maybe I'll forgo a little bit of it."* P7 valued humans' distinct ability and sense of responsibility to verify and screen for potential errors. However, she did not anticipate such ability and accountability to be implemented into AI, though alluding to a preference that AI be equipped with such accountability.

4.1.4 The expectation of AI to be unadaptable to new and changing circumstances made participants feel it is unuseful to trust or distrust AI. As discussed earlier, participants held the expectation that AI's advantage over that of humans' in terms of accuracy and flawlessness is confined to particular tasks: those that are straightforward, repetitive, and demand rapid computation. However, when it comes to more intricate tasks involving risk and uncertainty, participants believe that AI's impressive computing power might not suffice to adapt to new or changing situations. For instance, P7 (woman, HHA-T, M1) noted that, *"In simple task like this, I probably would trust them (DEMPC) more (if were AI). But maybe in more complicated tasks, things that requires maybe judging of an aircraft is going to crash, that's probably going to be more of a human person."* Similarly, P2 (woman, HHA-T, M1) explained how her comparison of trust in a human versus an AI teammate would depend on the type and complexity of the task, *"I guess in terms of just giving me the correct information, I may even trust AI more because it would be less error prone. But if we would face some uncertainty, I think I would trust more humans."* As these quote indicate, even though AI is expected to be better than humans with respect to fast and accurate computing capabilities, participants would still desire human intervention in the face of uncertainty and risk. Uncertainty and risk present unexpected situations that is likely beyond the AI's predefined parameters and rules. These new and changing circumstances can pose considerable challenges for AI systems to effectively navigate and address, especially without prior training or exposure to such scenarios.

Further, several participants mentioned that AI lacked human instinct, and the ability to learn from experience and self-improve to be adaptable to new circumstances and contextual needs. P7 (woman, HHA-T, M1) elaborated on this advantage that humans possess over AI, *"As of now, the way I think of AI is it still relies a lot on information, it has to get exact information to be able to get something. As a human, we possess another thing we call instinct. Once you do a certain task so many times I feel like humans develop this instinct of what exactly is the best course of action. New situations arise, that we don't know if the AI will be able to respond to, but human knows how to respond to things that might have never happened before based on a collections of experience of irrelevant things. The instinct part of being so experienced in a task is one thing that AI cannot yet substitute. Other than that, if you gave me absolutely enough data, AI would be the way to go."* P7's account elucidates one of the reasons why AI is expected to be unable to adapt to changing situations like humans do, by reflecting on the current limitations of AI in comparison to human cognition. Specifically, the irreplaceable quality of human instinct derived from accumulated experiences and intuitive decision-making is not as straightforward to be programmed into AI systems. P23 (man,

HHA-D, M1) also echoed this expectation, *"With human-human teams, we're not just machines, where maybe we see a necessity to do things differently, to deal with issues that come up, that maybe this program has never encountered before. That's not the case of the machine."* These perspectives and expectations have shaped participants initial decision to trust or distrust AI and human teammates before entering into the collaboration.

Many participants reported that it is unuseful to trust or distrust AI, because it would not improve and change its behavior in response to humans' trust or distrust. As P1 (man, HHA-D, M1) explained, *"Not trusting the AI doesn't necessarily mean the AI is bad, because they just do whatever it's programmed to do. So trusting and distrusting an AI isn't really very useful. But human human team, you definitely have more of a reason to need to trust your team members."* P1 at first reported that the AI's inability to adapt caused him to put less trust in AI than human, but then clarified that he felt the need for trusting a human teammate to be more meaningful than for trusting an AI, given the humans' adaptability and flexibility in response to changing circumstances including teammates' attitudes. P23 (man, HHA-D, M1) further explained that while humans can self-improve after experiencing teammates' distrust, he was unsure if AI could do the same. *"When there's distrust in humans, humans can learn from those mistakes. If you see distrust in yourself, you don't want to disappoint your teammates, and could possibly teach yourself something for next time. I'm not sure if you can do that with AI, if AI can teach themselves anything, or it still depends on us to teach them stuff."*

To summarize, our analysis reveals that prior to human-AI teaming, participants have held expectations of AI regarding its advantages over humans in terms of ability (less error-prone), benevolence (lower tendency to have hidden intention and conflicting interests), and integrity (less likely to be biased); as well as disadvantages in terms of error identification and correction, and adaptability. These expectations have shaped the initial trust they formed for human and AI in nuanced ways.

4.2 The Development of Situational Trust in the Process of Human-AI Teaming

Drawing on participants' reflections, we identify three ways situational trust was developed and maintained in the process of their collaboration in a HAT. Specifically, we illustrate how 1) verbal and behavioral responsiveness, 2) communication proactivity, and 3) the rectification and acknowledgement of mistakes fostered and/or rebuilt trust over the course of interactions.

4.2.1 Verbal and behavioral responsiveness bred situational trust. Reflecting on their collaborations, almost all participants mentioned that being responsive was a key factor for their trust in both human and AI teammates during their interaction. Our in-depth analysis revealed that such responsiveness encompassed 1) the promptness, timeliness, and consistency of replies, 2) the acknowledgement of the receipt and understanding of messages, as well as 3) behavioral responsiveness where teammates acted upon other team members' feedback.

The promptness, timeliness, and consistency of replies. Many participants reported to have based their trust on their teammates' "fast response" (e.g., P15, HHA-D, M5), "speed of communication" (e.g., P2, woman, HHA-T, M5), "instant feedback" (e.g., P17, man, HAA-D, M3), and "timely response" (e.g., P33, man, control, M3), as the promptness and timeliness demonstrated the teammates' "sense of responsibility" (e.g., P7, woman, HHA-T, M5). For instance, P35 (woman, HAA-T, M5) reported that *"proper communication and providing me information at the right time without taking too much time to give it to me, are the major factors which allowed me to gain some trust."* Additionally, some participants were particularly impressed by the AI teammate's consistency in being responsive, which could potentially be an advantage of AI over humans, as humans could get annoyed and impatient for being requested the same information over and over. P7 (woman,

HHA-T, M5) recalled, *"I keep asking him (AVO) for the radius, even though I think, through half of the first mission, it was always five. Other people or humans could have easily just skipped over or not answered, but he kept entering. I mean, it showed because at the end of the mission, it did change into a 2.5. I always want to make sure it's the correct information so I keep asking. And he still responded very consistently and very quickly. So that's the only factor because that's the only way we interact with it. So that's how I think he is trustworthy."* P7 believed that AVO's quality of being consistently responsive contributed to his trust in it. He considered such quality of AI to be an advantage because humans would not have been so patient with the repeated requests.

Responsiveness was also reflected in **the acknowledgement of the receipt and understanding of messages**, much like backchanneling [19], a common behavior in human conversations. However, the AI teammates in our study were not equipped with such abilities, which prompted some participants to have lowered their trust due to the AI's lack of backchanneling. As P45 (man, HHA-D, M3) noted, he trusted the human teammate more than the AI for his backchanneling behavior that allowed him to believe his message was understood, *"Whenever I was mentioning to AVO that we are going on a wrong track, it was not responding. Like he's not understanding what I'm telling. But DEMPC was giving response accordingly. So I get DEMPC is understanding me. So my trust level was building on DEMPC."* P20 (man, HHA-D, M5) also noted that such an acknowledgement also helped breed trust because it showed the teammate's caring of the joint endeavor. *"(to help me trust AVO) some acknowledgement will be great. I prefer to have at least a little bit of acknowledgement, just to know that they got my message, and it sometimes show that they care in a way, about what we're doing."* Indeed, these quotes have demonstrated the importance of backchanneling in cultivating trust during human-AI collaboration, partly through ensuring participants of other members' engagement in the shared task.

A lack of prompt response or backchanneling had led participants to develop various suspections regarding their teammates' ability, intention, and socio-emotional status. For instance, P13 (woman, HAA-D, M5) reported that a delayed response from both AI teammates had caused her to suspect something went wrong technically, *"Communication played a huge role (in building trust) because I'd based on how long they took to respond. When they took longer to respond, it made me feel they were unsure of themselves, maybe they're having technical issues on their side."* P17 (man, HAA-D, M5) suspected that delayed responses from the AI teammate was a sign of its being annoyed by his endless requests. *"Towards the end, the responses were delayed when I requested information. In the beginning, it was almost instantaneous. And I felt towards the end, it was like, am I kind of being resented for asking too many questions? The lack of fast response from the DEMPC, that added a little bit of distrust over the course."* Evidently, the simple act of acknowledging the receipt and/or understanding of humans' message (aka backchanneling) can reassure humans about the AI's competence, consistency, and active participation in the collaborative task, thereby sustaining humans' trust. Importantly, while in HHA teams participants could still trust the team based on the human teammate's backchanneling, in HAA teams, the complaints about the lack of responsiveness seemed prominent, to the extent that some participants even reported a feeling of marginalization within the human-AI team where the human was in the minority. For example, P9 (woman, HAA-T, M5) recalled that she felt being alienated when both of her AI teammates were not responsive to her verbally, even though they seemed to have responded to her request behaviorally. *"I constantly had to ask and they didn't respond to it, even though they changed the speed. And I just felt more alienated. I would just feel like I'm the only one that's talking and no one else is trying to engage, because maybe they have their own little way of communication."* For P9, the lack of backchanneling from AI teammates made her speculate that AI might use a distinct communication method apart from the conventional human text chat, which she lacked access to. Consequently, this led to her feeling estranged from the team of AI agents.

Behavioral responsiveness. Besides prompt verbal responses, participants' trust also hinged on the teammate's responsiveness reflected in their actions, particularly when they requested more than information, expecting the teammate to take specific actions. Quite a few participants expressed that they trusted the AI teammate because it *"listened to (them)"* (e.g., P41, man, HHA-T, M3) and *"followed command"* (e.g., P28, man, HAA-D, M3). As P32 (man, HHA-D, M3) said, *"Whenever I asked, it (AVO) would respond, and it would change it. So it felt trustworthy. And it was quite reactive to whatever solutions I was giving it."* Similarly, P23 (man, HHA-D, M3) also noted that his trust was mainly based on the AI's *"changing what I needed them to change."* It is important to note that P23 was the person who believed prior to collaboration (in Mission 1) that it was unuseful to trust or distrust AI due to its inability to adapt around humans' request. It appeared that the AI teammate's behavioral responsiveness had revised his prior expectations of AI to be unadaptable and fixed, thereby revised his initial trust.

When behavioral responsiveness did not match verbal responsiveness, participants' trust would suffer. For example, P3 (woman, HAA-T, M3) believed that her AI teammates once failed to take actions upon what they had said they would, which caused her trust to diminish. *"Something that plays into the trust side, is that they responded fast. But they didn't always do what they responded fast with, they didn't follow through for what they said, which lowered my trust a little bit."* Evidently, behavioral responsiveness should align with verbal responsiveness, and both are equally important. If there's a disconnect between what is promised and what is done, it can lead to confusion, doubt, or a lack of confidence in the AI's reliability or capabilities. Therefore, ensuring that verbal responsiveness is reflected through corresponding actions is crucial for trust development during human-AI teaming.

4.2.2 Communication proactivity bred situational trust. A second way to promote situational trust during human-AI teaming was proactive communication. This proactive communication involved providing the participants with the information they needed before they requested it (aka anticipatory information pushing), and periodically checking in with participants to keep them updated on the shared task and situation.

Anticipatory information pushing. The interdependent nature of the simulation task required team members to rely on one another's information to complete their own task. As such, providing the information one needed without their requesting it could make the communication more efficient and expedite the mission progression. For instance, P15 (man, HHA-D, M5) stated that his trust in the DEMPC was built upon how the DEMPC's proactive communication *"cut (his) job short"*, *"Communication was very key. I just told him (DEMPC), give me the radius. And then after that, I didn't even have to ask for the radius. He already gave it to me. And then we made the team work more efficiently. He raised my trust, just from communication alone."* Such proactive communication required that the teammate anticipated other members' information need once the team established a task norm. Failing to proactively fulfil the participants' information need when they were expected to through repeated interactions, could lead participants to lose trust. P4 (non-binary, HAA-D, M3) was not content with the AI teammates' lack of proactivity at Mission 3. *"(Trust was low) partly because of the lack of volunteering pertinent information. As someone used to working in human teams, I would be very upset if a human was planning something I was on, and did not give me the pertinent information that they knew I was going to need every single time. So having to ask them the exact same question every single mission that kind of contributed to lower the trust. I think it's such a basic, consistent need. And it's not like oh, I need to know this random thing this one time. It's every single mission, my job I need that piece of information."* P4 pointed out that due to the consistent nature of the information required, it should have been simple (*"not a fancy feature"*) to program anticipatory information pushing into the AI that was designed specifically for this task.

The AI's inability to consistently deliver the needed information without prompting was seen as a flaw in its design. As such, this lack of proactivity violated P4's (and many others') expectation of AI to exhibit fewer errors, thereby lowered their initial trust as the collaboration went.

Participants also emphasized the importance of **volunteering information to keep teammates in the loop with respect to the shared task situation**, as a way of being proactive in communication. P32 (man, HHA-D, M3) reported to have trusted the AVO because it proactively communicated its intended destination ahead of time, allowing him to prepare configurations accordingly. Furthermore, the AVO's regularly checking in with him to inquire if any adjustments were necessary also contributed to his trust and a successful collaboration. *"Whenever it (the AVO) was going to the next target, it was communicating it to me well in advance, so I had time to change my settings. And before reaching the target area and within the radius, it would ask me whether the altitude is proper or not. So I would say it was trustworthy."* P47 (woman, HAA-D, M5) reflected on the DEMPC's spread of distrust about the AVO and interpreted it as the DEMPC's proactively informing her of the shared task situation. *"(I trusted the DEMPC) because if there is anything wrong going on in the pipeline, like if AVO is on the wrong route, or it's taking the wrong directions, it is giving me up-to-date information."* Unlike anticipatory information pushing discussed previously, in these instances, the information provided by the teammate was not directly essential for the participants to perform their own task. Yet, receiving this information heightened the participants' overall awareness of the collaborative environment, which is often challenging to monitor within a team comprising multiple members. This increased awareness seemed to have allowed them to make informed decisions, enhance communication, and make necessary adjustments more effectively. More importantly, such proactive communication was regarded as a reflection of the teammate's sense of responsibility. P7 (woman, HHA-T, M5) summarized that her primary criterion for trusting a teammate and a team mainly rested on their display of responsibility, particularly evident through proactive communication. *"The attitude of the team is one big thing (to determine trust). You need to feel that they are responsible, that might come through their language, their proactivity. How do they follow up? Do they let you know what's going on? They update you on things."*

Our analysis suggests that despite both being a form of communication proactivity, "anticipatory information pushing" was regarded by participants as a must in human-AI teaming, whereas "volunteering information to keep teammates in the loop" was a plus.

4.2.3 The rectification and acknowledgement of mistakes contributed to (re)building situational trust. During human-AI teaming, participants experienced occurrences of AVO making mistakes at times as experiment manipulations. Obviously, mistakes can erode trust, particularly when it comes to an AI teammate that is expected to be nearly flawless. The question then arises: to what extent does correcting a mistake contribute to the restoration of trust? Our study suggested that the rectification or simply the acknowledgement of mistakes can increase participants' trust. Interestingly, for some people, this acknowledgment and correction can even lead to higher levels of trust in a teammate compared to a scenario where the teammate made no mistakes. As P10 (man, HAA-D, M3) noted, *"Mistakes are fine. The fact that AVO was correcting its mistake made me trust him even more (than in M1 when AVO made no mistakes)."* Collaborating with an error-correcting AI teammate had violated some individuals' prior expectations of AI in a positive way. P19 (man, HHA-T, M3) explained that the AI's ability to correct its mistake had compensated for its failure to meet his expectation of perfection. *"While I did say (in M1 interview) AI should be perfect, but seeing that it could correct its error made me confident in the program. Minor errors, not critical, not affecting other's work, that would be fine."* Like P19, many participants expressed leniency towards the AI's minor and corrected mistakes, contrasting their firm belief before interaction that AI should be flawless.

Trust can be updated moment by moment as participants accumulated evidence of whether the teammates would voluntarily realize and correct their mistakes. Drawing on participants' reflections, it became evident that their trust in the teammates was not constant, but fluctuated in response to their observation of the teammates' ability to identify and correct errors, which was primarily exhibited through communication. For example, P10 (man, HAA-D, M5) recalled, *"On the second mission, he (AVO) made a few mistakes and then he didn't really correct himself so I didn't really trust him at that time because I had to change the record settings myself. But Mission four and five. He gave wrong info, but he corrected the info here. So I started building more trust."* Through collaboration, some were able to affirm that the teammates would "always" rectify their mistake when they occasionally made one. This experience had led to a relatively stable trust, contingent on the teammates maintaining consistent behavior. As P13 (woman, HAA-D, M5) said, *"When it would give me incorrect information, it would always make sure to correct it right afterwards. So then I would know that even if they gave me the wrong answer, I know I could still trust them to an extent because at the end of the day, they still give me the proper answer that I would need to complete the mission."*

If an AI teammate cannot realize and correct its mistake, they had better be able to correct upon request, or at least acknowledge the mistake when others pointed out. While not as preferable as voluntarily identifying and rectifying its own mistake, acknowledging it can also demonstrate a sense of responsibility, thereby maintaining trust. For instance, P36 (woman, HAA-D, M3) explained that *"I'm not distrustful of it because of those mistakes. They correct themselves as soon as I pointed out so I still trusted them. If I pointed out and they ignore my message, or still insist on their previous incorrect information, then there's distrust, but I didn't see this circumstance."* In some cases where participants failed to get the AVO to correct its error (due to the participants' failing to use the restrictive language comprehensible by the AI), participants expressed frustration and significant loss of trust. P1 (man, HHA-D, M3) noted that acknowledgment of mistakes was a manifestation of humanness, which the AI teammate lacked, *"The distrust was mainly that they (AVO) didn't actually own up to their mistakes. They just weren't acting like a human. Even if you make mistakes, that's okay, because by acknowledging it, it means you're all trusting each other to correct your mistakes. It should communicate its errors to be more trustworthy, by showing vulnerabilities in front of your teammates, that's the humanness."* For P1, acknowledging mistakes has the potential to foster mutual trust by signaling that one is willing to be vulnerable and entrusts their teammates with that vulnerability.

4.3 The Impact of (Dis)Trust Learned through Human-AI Teaming on Individuals' Subsequent Expectation and Collaboration within a HAT

4.3.1 Learned (dis)trust in one AI tended to be carried over to another AI. Our study revealed that unlike the process of establishing trust with human teammates, which is based on individual cases and involves participants refraining from making judgments before interacting, participants tended to generalize their trust in AI, carrying their trust or distrust established from previous encounters over to the current one. For instance, having interacted with an error-prone AI could lead participants to distrust AI in general.

We asked participants in HHA conditions whether and how their trust would change if DEMPC (played by a human) were an AI. Many who had experienced glitches with the AI expressed that they would naturally distrust DEMPC. As P20 (man, HHA-D, M5) reported, *"Based upon my experience, if DEMPC also happens to be an AI, I would have more distrust with him. After going through how the AVO is an AI and how there's been some couple of hiccups here and there. I'd say, it'd be better if it was not an AI."* Those who had relatively smooth experience with the AI, or had not yet encountered issues tended to trust DEMPC if it were an AI. P26 (woman, HHA-T, M3) noted

that *"if they (DEMPC) were an AI, I feel like I would trust them more to perform the mission, because the AI has done a good job up to this point. So I would have confidence in their program."* Indeed, like P20 and P26, many participants appeared to have formed expectations regarding the performance and capabilities of AI for future encounters or hypothetical scenarios based on their experience with AI in the just-finished episode. Their trust or distrust formed in the recent interaction shaped or even determined their trust or distrust in the AI teammate in upcoming collaborations or even beyond the current context.

In contrast, participants in HAA conditions avoided making premature judgments about whether they would trust DEMPC if it were a human teammate. For instance, P13 (woman, HAA-D, M5) explained that she could not decide her trust because *"it depends on who the person is. People are more likely to trust someone that they know as opposed to a stranger... Even with the same person, there's a wide range, because you can't guarantee that person is as focused as he or she previously (was). So you really need to work with them and see... But from what I've seen people react to AI, a lot of people will just trust AI right from the get-go without having any experience with it."* When it comes to human teammates, participants like P13 tended to withhold trust judgments. Not a single participant had indicated that they would extend their trust from one human to another.

Participants' tendency to generalize their trust or distrust in AI potentially results from the perception that AI is fixed and unadaptable. They expected that individual AI agents in the same task scenario, despite occupying distinct roles, to be similarly programmed and therefore equally trustworthy or untrustworthy. For example, P7 (woman, HHA-T, M5) compared the influence of learned distrust on one's inclination to trust AI versus human teammates in subsequent interactions, and reasoned that *"AI teammates, if it makes a wrong decision once, everybody's gonna think that something is wrong with the program. And that distrust is harder to rectify, as in there must be something wrong with the program itself. If it's for a human, it might have been that person is probably tired that day. But it's not the same case with an AI. If there's a mistake, if you copy and paste the program into a different computer, it will behave the exact same way."* Indeed, unlike humans, where an occasional mistake might be attributed to emotional or physiological factors, an AI's error tends to be perceived as a fundamental flaw within the AI itself. The lack of contextual factors influencing AI behavior, such as emotions or situational variability, contributes to the perception that if the program behaves incorrectly once, it will likely repeat the same mistake consistently across different instances or platforms.

4.3.2 Learned distrust in AI led to effortful monitoring behaviors. After experiencing occurrences of AVO's mistakes, many participants reported to have become more "wary" (e.g., P26, woman, HHA-T, M5), "vigilant" (e.g., P32, man, HHA-D, M3), and "cautious" (e.g., P7, woman, HHA-T, M5). They expressed frustrations of having to "keep an eye out" (e.g., P26, woman, HHA-T, M5) for future mistakes, and having to "overcompensate" (e.g., P9, woman, HAA-T, M5) by allocating some of their mental capacity from performing their own tasks to monitoring the AI teammate. For instance, P8 (woman, HHS-D, M5) complained that AVO repeatedly messing up the target waypoints *"completely destroyed my trust. Every time I have to scroll up to see if AVO was going back to the area that we went through already. That's why I was saying I really got flustered, is because my attention was drawn elsewhere. I could barely do my own settings."* For P8, performing the task and coordinating with the other two teammates was already overwhelming. After experiencing AVO's mistakes in Mission 2, she reported to have had to monitor AVO's trajectory every time in case it made mistakes again that would jeopardize the team's effort. Such monitoring behavior further burdened her.

In HAA conditions, some even reported to have monitored DEMPC who never made mistakes, reasoning that since AVO was unreliable, DEMPC's reliability could not be guaranteed given that AVO and DEMPC must have been created by the same programmer. For instance, P9 (woman,

HAA-T, M5) reported that *"Later on, I was double checking DEMPC's information as well. Knowing that AVO wasn't as reliable, I was thinking maybe at some point DEMPC would make a mistake, because they're based on the same program. I don't want them to mess me up. So I'm going to double check no matter what."*. Interestingly, in HAA conditions, DEMPC's unwarranted spread of (dis)trust about AVO contributed to another reason why participants expressed a desire to monitor both AI teammates. As P16 (man, HAA-T, M3) explained, *"As soon as I noticed that it (DEMPC) was saying good messages about the AVO, it just made me more aware that they might make mistakes. My lingering distrust applied to them both. I was on the lookout. I would look more closely at their messages. And I'll make sure they follow through, just double check that he was actually doing that."* P9 (woman, HAA-T, M5) also echoed this by saying, *"The DEMPC, I was on the fence, I'm like, 'I'm watching you. You haven't done anything wrong. But you're agreeing the thing that has done things wrong. I'm gonna double check to make sure that you're doing your correct job, too."* These accounts have further demonstrated that one traumatic experience with one AI can leave a lasting impact, making participants overly cautious in subsequent interactions and even monitor another AI that is not necessarily bad. Interestingly, in HHA conditions, when it was a human teammate who spread undeserved (dis)trust about AI, there was no mentioning of a desire to monitor the human teammate.

In summary, Figure 3 synthesizes our findings, which illustrates the diverse factors influencing trust across different phases of human-AI teaming, and the interplay among them. Specifically, the blue box summarizes the influence of prior expectations of AI versus human teammates on the initial trust formed before human-AI teaming. The orange box lists the communication behaviors that can foster situational trust in the process of human-AI teaming. The green box pinpoints the consequences of learned trust on subsequent expectation of AI's ability and collaboration with it. In this diagram, "current interaction episode" refers to the most recent interaction episode participants reflected upon, where they experienced a cycle of the initial trust (formed before collaboration) being strengthened, maintained, or modified during collaboration to form situational trust, which in turn evolves into learned trust, which subsequently influences a new round of initial trust formed before "future collaboration episodes" in the same or different context.

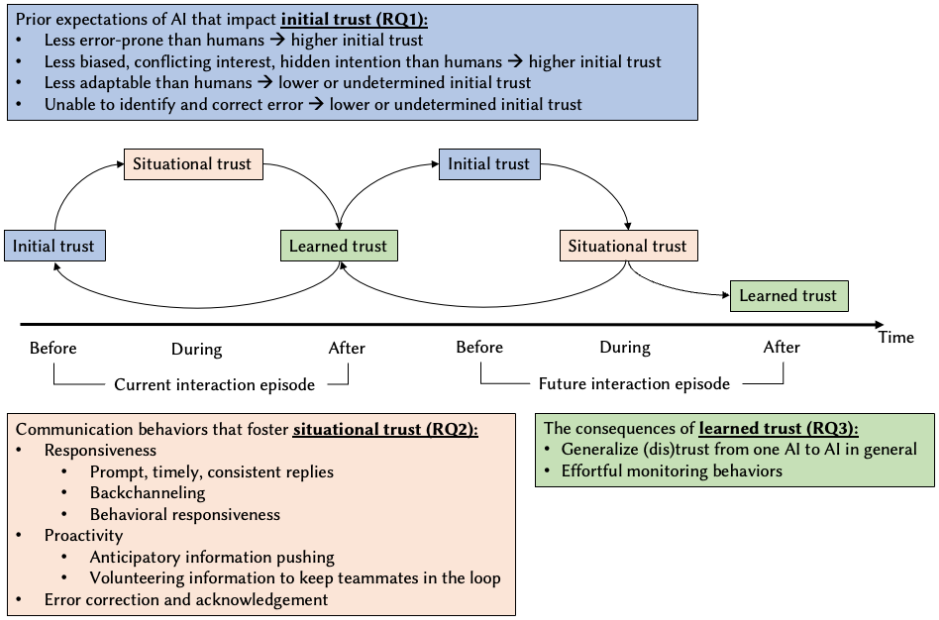

Fig. 3. Trust Evolvement Over Time in a HAT

5 DISCUSSION

5.1 Explaining Temporal Variations of Trust for AI and Human Teammates

Our study reveals that prior to human-AI teaming, individuals held distinct **initial levels of trust** (**RQ1**) in their human and AI teammates, primarily due to varying expectations regarding their capability, integrity, benevolence, and adaptability. This trust divergence between human and AI teammates also existed in the consequences of **learned trust** (**RQ3**), where trust learned through human-AI collaborations was generalized to the broader category of AI but not extended to humans in general. However, the ways in which **situational trust** (**RQ2**) during human-AI teaming was fostered were analogous and applicable to both human and AI teammates. We discuss the implications of these findings in light of extant literature.

5.1.1 Effective calibration of trust in AI teammates prior to interaction should go beyond reliability and encompass affective dimensions of trust. Findings of our study suggest that most participants held high and unrealistic expectations of AI, resulting in their higher initial trust in AI than human teammates. This echoes recent explorations that AI is expected to be an ideal version of a human [113] and more trusted than humans [110]. Our work further specifies that the elevated expectations center around the AI's presumed accuracy (i.e., less prone to errors), integrity (i.e., less biased), and benevolence (i.e., less likely to have hidden intention), which may need to be calibrated down to align with the actual characteristics of the AI teammate. Conversely, alongside the high expectations placed on AI, there are also instances of low expectations regarding the AI's adaptability and error correction ability, which may need to be calibrated upwards. While the prevailing body of research primarily focuses on calibrating trust in AI solely based on its reliability (e.g., [7, 85, 106]), our study highlights the necessity to also calibrate trust concerning the AI's integrity, benevolence, and adaptability, particularly in team contexts. Our work expands the current emphasis on trust

calibration for AI systems [106] not only by identifying additional aspects for trust calibration, but also by showing that distinct facets of trust require varying calibration approaches.

5.1.2 Situational trust is primarily fostered through a wide range of communication behaviors. Despite the growing awareness that communication is central to the maintenance and breeding of trust in human-AI teaming [29, 74, 82], it is yet unclear what trust-breeding communication entails. Our study essentially provides a typology of communication behaviors and their functions in fostering situational trust over the course of human-AI teaming and collaborations, which applies for both human and AI teammates.

First, our work identifies two components of proactive communication –anticipatory information pushing, and volunteering information to keep teammates in the loop. This corroborates and extends the recent work [111] that has demonstrated the importance of proactive communication in the development of trust in dyadic HATs in multi-player gaming context. Our work highlights that the first component is a must whereas the latter is a plus. Because the latter behavior goes beyond one's designated duties in the team, highlighting a strong sense of responsibility highly esteemed in teamwork. This attribute holds particular significance in distributed multi-member teams, where members encounter greater challenges than those in dyadic teams in terms of sustaining shared situation awareness [90], and mitigating the potential for miscommunication [19].

Second, we also identified three components of responsiveness that goes beyond simply quick responses that previous studies [49, 111] have suggested: namely, the promptness, timeliness, and consistency of replies; the acknowledgement of the receipt and understanding of messages (or backchanneling); and behavioral responsiveness. In particular, backchanneling has often been overlooked or deemed trivial in designing AI systems [48], partly due to their lack of direct relevance to task execution that AI is primarily designed for. However, contrary to this perspective, backchanneling serves a vital role in facilitating task coordination by signaling the listener's attention and active engagement [19]. Instead of being extraneous to the task, backchanneling aids in establishing effective communication dynamics, thereby enhancing task-related coordination and efficiency within human-AI interactions. Additionally, a recent study [6] has found that AI's backchanneling behavior can influence humans' perceptions of its personality. As human-AI collaboration becomes increasingly relying on natural language communication, principles of human conversation norms such as incorporation of backchannels, should be accounted for in task-oriented AI design.

Third, the acknowledgment and rectification of mistakes during human-AI teaming is also done through communication. This is similar to the well-studied concept of trust repair [53, 55, 93]. One common repair strategy is apology. There are discrepancies in the effects of apology on trust. For instance, while [93] found no significant effect of apologies versus denial on trust, [53] suggested that apologies served to regain trust and were better than denying an error. Our work lends insights into such discrepancies by shedding light on the role of the consequences caused by AI errors. In our study, participants highlighted that they found it most desirable when the AI teammate could identify and rectify an error autonomously, even before other team members became aware of it, thus avoiding inconvenience to others. Importantly, verbal acknowledgement of such realization and correction is what really boosted trust. If this is impossible at the moment, correcting the error upon request is second desirable. At the very least, the AI teammate should acknowledge its error verbally. While a few mentioned that an apology would work, most participants did not demand the form of error correction or acknowledgement.

5.1.3 Mitigating the carryover effect of learned trust. Our study suggests that having collaborated with AI teammate(s), individuals will generalize their trust or distrust in one AI teammate to all AI. This is in line with [44]'s finding that the failure of the subsystem can result in a decrease in

trust toward the entire system. Similarly, [102] found that when one system exhibited inaccuracy, it prompted more verifications and subsequently lowered the subjective trust placed in similar but independent systems. Whether it is trust or distrust, the carryover effect will be taken to subsequent episodes of interaction and impact the individual's expectations of AI, thereby influencing the initial trust formed for the new episode. This cycle of trust development is illustrated in Figure 3. One of the reasons why the problem of over-reliance does not apply to that between humans is that people do not generalize trustworthiness from one individual to another. The carryover effect of learned trust can cause problems of miscalibration [106] of trust, causing people over-trust or under-trust an AI teammate in subsequent interactions or in a different context, resulting in either blindly, wrongfully accepting AI's decisions/suggestions, complacency [97], or wrongfully rejecting or disregarding them [28].

5.2 Implications for Methodology and Theory for Future Human-AI Collaboration Research

Our work has demonstrated that different factors and sub-constructs of trust receive varying emphasis at different phases of human-AI teaming. This suggests that measuring trust at the conclusion of an interaction episode will not likely capture the intricacies of trust fluctuation and the interplay of team dynamics. This suggests the need for trust measures to account for the temporal variations of trust, which could be achieved through several methodological innovations.

First and foremost, the easiest to implement is by adding more time points to measure trust prior, during, and after each interaction. For instance, prior to human-AI collaboration, individuals' initial trust should be collected to serve as the baseline. [38]'s dynamical resilience metrics provides a useful tool for measuring team team resilience in dynamic socio-technical environments. Informed by our study, it is important to note that individuals' propensity to trust is different from their propensity to trust an AI teammate. Importantly, the sub-constructs of these pre-, during-, and post-trust measures must be examined separately as the aspects of trust valued at each stage will likely be attributed different weights.

Second, during human-AI teaming, even a short episode of interaction can involve many subtle cues (e.g., a delay in response) that can affect individual's trust perceptions. To capture the nuances, trust can be measured dynamically even during interaction in an unobtrusive manner. For instance, future research could leverage the system itself to detect behavioral trust through indicators of monitoring behavior [23]. Psychophysiological equipment such as eye-tracking equipment could also be used to indicate trust and its fluctuation across time, as gaze behavior has been used as an indicator of trust in automated driving [42].

Third, the impact of learned trust on subsequent perceptions of and collaborations with AI suggests a need for longitudinal studies that enable researchers to capture trust dynamics across different contexts, and identify patterns that might not be apparent in one short-term experiment. Fourth, to gain a holistic understanding of the wax and wanes of trust within HAT context and its influencers, field research [79, 84] with individuals who actually collaborate with AI on a daily basis is the best, as it offers opportunities to identify real-world challenges and constraints with respect to both technology and team dynamics, which may lead to discovery of unanticipated phenomena and considerations of trust. Alternatively, experience sampling [101] offers a useful tool to capture the moment-by-moment changes of trust and can be implemented to investigate the day-to-day experience of multi-player gamers for whom HAT is not an experimental concept but their everyday reality [113]. In sum, our study reveals the temporal and dynamic nature of trust in HATs which calls for innovative or a combination of methodologies to capture.

Our study also shed light on the directions of theoretical advancements of trust within HATs. HAT research on trust has benefited from theories of organizational and interpersonal trust [69, 71] and

theory of trust in automation [60]. However, our work suggests that humans' trust in AI teammate is not entirely the same as that in human teammates especially before and after teaming, suggesting that organizational and interpersonal trust theories may not adequately explain trust within HATs. Additionally, the team context also relegate the theory of trust in automation inadequate for understanding trust in HATs. To date, efforts to theorize trust development within HATs are rare, despite a few exceptions that draw heavily on trust in automation frameworks [7, 10, 100]. Our work provides a starting point to inductively build theory of trust within HATs from a temporal perspective. Specifically, our findings elucidate the feedback loops between trust, expectations and behaviors, illuminating how changes in trust influence subsequent actions and vice versa over time.

5.3 Designing Multi-faceted Multi-Phase and Dynamic Trust Calibration for HATs

Grounded in our findings that reveal the temporal variations of trust, its sub-dimensions, and related constructs at different stages of human-AI teaming, we provide several design suggestions for effective human-AI teaming to maintain an appropriate level of trust. By outlining the interplay between **what to calibrate, when to calibrate, and how to calibrate trust**, our suggestions revolve around trust calibration to be **multi-faceted, multi-phase, and dynamically adaptable** to contextual needs.

First, trust is a multi-faceted concept that encompasses the perceived trustworthiness of the trustee, which further consists of the trustee's ability, integrity and benevolence [69]. However, as shown earlier, a vast majority of AI trust calibration strategies have focused on AI's reliability [106] or do not differentiate these sub-constructs [106], despite a few exceptions (e.g., [32]) Our work has demonstrated that besides AI's reliability, its integrity, benevolence, and adaptability aspects also warrant calibration. Indeed, the presence of algorithmic bias in [58, 59] calls for humans' expectations of AI's integrity to be wisely calibrated. To avoid unwarranted high trust and heuristic thinking [9], the human-AI teaming system might consider integrating uncertainty cues regarding the integrity principles that the AI is supposed to adhere to. This approach parallels the ways in which uncertainty cues of AI reliability have been adopted in most trust calibration research [57, 114]. The integrity principles could be presented as a fairness checklist as suggested by [47, 66]. Our work also highlights the need to calibrate humans' expectations about AI's adaptability. With increasing focus on designing AI to be more adaptable [36], humans' AI literacy should also be updated to incorporate whether and how an AI agent can be adaptable in the collaborative environment.

Second, our work suggests that interventions of trust calibration implemented at different phases of human-AI teaming should address different aspects of AI's perceived trustworthiness and should account for individual differences. For instance, before collaboration, calibration should target individuals' prior expectations through personalized training [107] where humans are educated about the actual ability, integrity, and adaptability of the AI to correct any misperceptions and be properly prepared for human-AI teaming. Trust calibration is more effective to be applied at early stages of interaction, when it is decisive for trust development [108]. During collaboration, AI behaviors can be designed to address the individual human teammate's potential over- or under-trust by purposefully violating his/her prior expectations. For instance, if the person indicates a low initial trust in AI based on the expectation of its being unable to adapt, the AI can display adaptive behaviors early on to reassure that person. For purpose of preventing applying it broadly to other situations, learned trust should also be calibrated following a human-AI teaming episode. To do this, a debrief session is needed to provide explanations about how specific incidents during the collaboration influenced the development or decay of trust, and emphasize that these instances are unique to the system and not necessarily reflective of future scenarios.

Third, trust calibration can be made dynamical and adapt to the situational needs of individual human teammates. For instance, the system could detect human team members' trust in AI team members through behavioral cues such as monitoring behaviors, and juxtapose such cues against the actual accuracy of AI to determine if the human is over-trusting, under-trusting or applying appropriate trust toward the AI teammate, thereby dynamically display calibration cues (along the lines of [16, 81]) appropriate at the moment. Our work identifies a typology of communication behaviors that foster trust during collaboration. Future human-AI teaming systems might consider incorporating this typology of behaviors to create a dynamic feedback feature, through which human team members could give real time feedback with regard to their communication needs from the AI. This could be easily implemented as a hovering menu with a set of request options: request response, proactivity, mistake correction, as well as a list of sub-requests under each. These requests can be configured to quantify the levels. For instance, humans could complain through the feedback feature about the AI's lack of immediate response, and request it to give faster and more frequent responses; or, if they feel the pushing of situation-related information is too detailed to be excessive [111], they could request it to be less frequent or less detailed.

5.4 Limitation and Future Work

Findings of this study need to be interpreted with several limitations in mind. First, participants' exposure to human-AI teaming occurred within an experimental setting, where deliberate manipulations created the main incidents (e.g., AI teammate making mistakes) that participants experienced and reflected upon. This controlled setup might have restricted the diversity of factors we could identify concerning participants' development of trust and distrust during human-AI teaming. Future research could validate these findings in more naturalistic environments such as conducting field research in the workplace or collaborative projects, to offer a broader spectrum of incidents and interactions. Second, while we purposefully chose the collaboration task and simulation environment to provide 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 may have been relatively focused. Due to its focus on task completion, 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 the temporal variations of trust within a HAT. Lastly, all the participants in our study were university students. Despite our best effort to recruit a more diverse sample encompassing various racial and ethnic backgrounds, the resulting sample consisted of an equal split between White and Asian individuals. Future research could focus on expanding the diversity of the sample, specifically targeting a broader range of educational backgrounds and a more extensive representation of racial and ethnic groups.

6 CONCLUSION

Trust is dynamic and changes moment by moment in response to team interactions. Despite the growing interest in researching trust in human-AI teaming, the temporal and dynamic nature of trust in HATs is still significantly understudied. To address this gap, we employed a multi-phase qualitative method to interview 45 individuals who collaborated in a three-member human-AI or human-only team at three time points over the course of a series of teaming episodes. Our study reveals that prior to human-AI teaming, individuals tended to have higher initial trust in AI than

human teammates due to different expectations regarding the AI and human's ability, integrity, benevolence, and adaptability. During human-AI teaming, these prior expectations could be affirmed or violated as individuals gain evidence through observation of the AI and human teammate's verbal and behavioral responsiveness, communication proactivity, and acknowledgement and rectification of mistakes. As such, the initial trust could be maintained, revised, and updated moment by moment as situational trust. Finally, our work showed that individuals' learned trust and distrust in AI teammate can be carried over to their subsequent expectations of and collaborations with AI in the same and different context. Our study advances the understanding of trust development in team contexts by identifying the temporal variations of trust. We also provide valuable insights into the effective calibration of trust for human-AI teams.

ACKNOWLEDGMENTS

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.

REFERENCES

[1] Ronald Arkin. 2009. *Governing lethal behavior in autonomous robots*. CRC press.
[2] Eugénie Avril. 2023. Providing different levels of accuracy about the reliability of automation to a human operator: impact on human performance. *Ergonomics* 66, 2 (2023), 217–226.
[3] Annette Baier. 1986. Trust and Antitrust. *Ethics* 96, 2 (1986), 231–260. http://www.jstor.org/stable/2381376
[4] Daniel Balliet and Paul AM Van Lange. 2013. Trust, conflict, and cooperation: a meta-analysis. *Psychological bulletin* 139, 5 (2013), 1090.
[5] Gagan Bansal, Alison Marie Smith-Renner, Zana Buçinca, Tongshuang Wu, Kenneth Holstein, Jessica Hullman, and Simone Stumpf. 2022. Workshop on Trust and Reliance in AI-Human Teams (TRAIT). In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–6. https://doi.org/10.1145/3491101.3503704
[6] Peter Blomsma, Gabriel Skantze, and Marc Swerts. 2022. Backchannel behavior influences the perceived personality of human and artificial communication partners. *Frontiers in Artificial Intelligence* 5 (2022), 835298.
[7] Philip Bobko, Leanne Hirshfield, Lucca Eloy, Cara Spencer, Emily Doherty, Jack Driscoll, and Hannah Obolsky. 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 (2023), 310–334.
[8] Christina Breuer, Joachim Hüffmeier, and Guido Hertel. 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 (2016), 1151.
[9] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.
[10] Sabrina Caldwell, Penny Sweetser, Nicholas O'Donnell, Matthew J Knight, Matthew Aitchison, Tom Gedeon, Daniel Johnson, Margot Brereton, Marcus Gallagher, and David Conroy. 2022. An agile new research framework for hybrid human-AI teaming: Trust, transparency, and transferability. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 12, 3 (2022), 1–36.
[11] Kendall Carmody, Cherrise Ficke, Daniel Nguyen, Arianna Addis, Summer Rebensky, and Meredith Carroll. 2022. A Qualitative Analysis of Trust Dynamics in Human-Agent Teams (HATs). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 66. SAGE Publications Sage CA: Los Angeles, CA, 152–156.
[12] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. sage.
[13] Jessie YC Chen. 2018. Human-autonomy teaming in military settings. *Theoretical issues in ergonomics science* 19, 3 (2018), 255–258.
[14] Jessie YC Chen and Michael J Barnes. 2014. Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems* 44, 1 (2014), 13–29.

[15] Jessie YC Chen, Michael J Barnes, Anthony R Selkowitz, Kimberly Stowers, Shan G Lakhmani, and Nicholas Kasdaglis. 2016. Human-autonomy teaming and agent transparency. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces*. 28–31.
[16] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with trust for human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. 307–315.
[17] Xusen Cheng, Guopeng Yin, Aida Azadegan, and Gwendolyn Kolfschoten. 2016. Trust evolvement in hybrid team collaboration: A longitudinal case study. *Group Decision and Negotiation* 25 (2016), 267–288.
[18] Li-Fang Chou, An-Chih Wang, Ting-Yu Wang, Min-Ping Huang, and Bor-Shiuan Cheng. 2008. Shared work values and team member effectiveness: The mediation of trustfulness and trustworthiness. *Human relations* 61, 12 (2008), 1713–1742.
[19] Herbert H Clark. 1996. *Using language*. Cambridge university press.
[20] Nancy J Cooke and Steven M Shope. 2004. Synthetic task environments for teams: CERTT's UAV-STE. In *Handbook of human factors and ergonomics methods*. CRC Press, 476–483.
[21] Ana Cristina Costa, C Ashley Fulmer, and Neil R Anderson. 2018. Trust in work teams: An integrative review, multilevel model, and future directions. *Journal of Organizational Behavior* 39, 2 (2018), 169–184.
[22] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz studies—why and how. *Knowledge-based systems* 6, 4 (1993), 258–266.
[23] Bart A De Jong and Kurt T Dirks. 2012. Beyond shared perceptions of trust and monitoring in teams: Implications of asymmetry and dissensus. *Journal of Applied Psychology* 97, 2 (2012), 391.
[24] Bart A De Jong, Kurt T Dirks, and Nicole Gillespie. 2016. Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of applied psychology* 101, 8 (2016), 1134.
[25] Bart A De Jong and Tom Elfring. 2010. How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Academy of Management journal* 53, 3 (2010), 535–549.
[26] Jaap J Dijkstra. 1999. User agreement with incorrect expert system advice. *Behaviour & Information Technology* 18, 6 (1999), 399–411.
[27] Theo Dimitrakos. 2002. A service-oriented trust management framework. In *Workshop on Deception, Fraud and Trust in Agent Societies*. Springer, 53–72.
[28] Stephen L Dorton, Samantha B Harper, and Kelly J Neville. 2022. Adaptations to trust incidents with artificial intelligence. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 66. SAGE Publications Sage CA: Los Angeles, CA, 95–99.
[29] Wen Duan, Nathan McNeese, and Rui Zhang. 2023. Communication in Human-AI Teaming. *Group Communication* (2023), 340–352.
[30] Mary T Dzindolet, Linda G Pierce, Hall P Beck, and Lloyd A Dawe. 2002. The perceived utility of human and automated aids in a visual detection task. *Human factors* 44, 1 (2002), 79–94.
[31] Mica R Endsley. 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 (2015), 1–34.
[32] Connor Esterwood and Lionel P Robert. 2021. Do you still trust me? human-robot trust repair strategies. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE, 183–188.
[33] Jing Feng, Joseph Sanchez, Robert Sall, Joseph B Lyons, and Chang S Nam. 2019. Emotional expressions facilitate human–human trust when using automation in high-risk situations. *Military Psychology* 31, 4 (2019), 292–305.
[34] Christopher Flathmann, Wen Duan, Nathan J Mcneese, Allyson Hauptman, and Rui Zhang. 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 (2024), 1–32.
[35] Fiona Fui-Hoon Nah, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. 2023. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. , 277–304 pages.
[36] Omid Gheibi, Danny Weyns, and Federico Quin. 2021. Applying machine learning in self-adaptive systems: A systematic literature review. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 15, 3 (2021), 1–37.
[37] Ella Glikson and Anita Williams Woolley. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* 14, 2 (2020), 627–660.
[38] David AP Grimm, Jamie C Gorman, Nancy J Cooke, Mustafa Demir, and Nathan J McNeese. 2023. Dynamical Measurement of Team Resilience. *Journal of Cognitive Engineering and Decision Making* 17, 4 (2023), 351–382.
[39] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors* 53, 5 (2011), 517–527.
[40] Allyson I. Hauptman, Wen Duan, and Nathan J. Mcneese. 2022. The Components of Trust for Collaborating With AI Colleagues. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, Taiwan) *(CSCW'22 Companion)*. Association for Computing Machinery, New York, NY,

USA, 72–75. https://doi.org/10.1145/3500868.3559450
[41] Allyson I Hauptman, Beau G Schelble, Wen Duan, Christopher Flathmann, and Nathan J McNeese. 2024. Understanding the influence of AI autonomy on AI explainability levels in human-AI teams using a mixed methods approach. *Cognition, Technology & Work* (2024), 1–21.
[42] Sebastian Hergeth, Lutz Lorenz, Roman Vilimek, and Josef F Krems. 2016. Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving. *Human factors* 58, 3 (2016), 509–519.
[43] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.
[44] Kai Holländer, Philipp Wintersberger, and Andreas Butz. 2019. Overtrust in external cues of automated vehicles: an experimental investigation. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*. 211–221.
[45] Lixiao Huang, Nancy J Cooke, Robert S Gutzwiller, Spring Berman, Erin K Chiou, Mustafa Demir, and Wenlong Zhang. 2021. Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In *Trust in human-robot interaction*. Elsevier, 301–319.
[46] Ming-Hui Huang and Roland T Rust. 2018. Artificial intelligence in service. *Journal of service research* 21, 2 (2018), 155–172.
[47] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 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* (Virtual Event, Canada) *(FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 624–635. https://doi.org/10.1145/3442188.3445923
[48] Vidit Jain, Maitree Leekha, Rajiv Ratn Shah, and Jainendra Shukla. 2021. Exploring semi-supervised learning for predicting listener backchannels. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.
[49] Sirkka L Jarvenpaa and Dorothy E Leidner. 1999. Communication and trust in global virtual teams. *Organization science* 10, 6 (1999), 791–815.
[50] Karen A. Jehn and Elizabeth A. Mannix. 2001. The Dynamic Nature of Conflict: A Longitudinal Study of Intragroup Conflict and Group Performance. *Academy of Management Journal* 44 (2001), 238–251. https://api.semanticscholar.org/CorpusID:17800484
[51] Noel D Johnson and Alexandra A Mislin. 2011. Trust games: A meta-analysis. *Journal of economic psychology* 32, 5 (2011), 865–889.
[52] Prasert Kanawattanachai and Youngjin Yoo. 2002. Dynamic nature of trust in virtual teams. *The Journal of Strategic Information Systems* 11, 3-4 (2002), 187–213.
[53] Spencer C Kohn, Daniel Quinn, Richard Pak, Ewart J De Visser, and Tyler H Shaw. 2018. Trust repair strategies with self-driving vehicles: An exploratory study. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 62. Sage Publications Sage CA: Los Angeles, CA, 1108–1112.
[54] Sherrie YX Komiak and Izak Benbasat. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly* (2006), 941–960.
[55] E. S. Kox, L. B. Siegling, and J. H. Kerstholt. 2022. Trust development in military and civilian human–agent teams: The effect of social-cognitive recovery strategies. *International Journal of Social Robotics* 14, 5 (July 2022), 1323–1338. https://doi.org/10.1007/s12369-022-00871-4
[56] Nicole C Krämer, Gale Lucas, Lea Schmitt, and Jonathan Gratch. 2018. Social snacking with a virtual agent–On the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *International Journal of Human-Computer Studies* 109 (2018), 112–121.
[57] Alexander Kunze, Stephen J Summerskill, Russell Marshall, and Ashleigh J Filtness. 2019. Automation transparency: implications of uncertainty communication for human-automation interaction and interfaces. *Ergonomics* 62, 3 (2019), 345–360.
[58] Susan Leavy. 2018. Gender bias in artificial intelligence: the need for diversity and gender theory in machine learning. In *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*. ACM, Gothenburg Sweden, 14–16. https://doi.org/10.1145/3195570.3195580
[59] Susan Leavy, Eugenia Siapera, and Barry O'Sullivan. 2021. Ethical Data Curation for AI: An Approach based on Feminist Epistemology and Critical Theories of Race. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Virtual Event USA, 695–703. https://doi.org/10.1145/3461702.3462598
[60] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.
[61] Roy J Lewicki, Daniel J McAllister, and Robert J Bies. 1998. Trust and distrust: New relationships and realities. *Academy of management Review* 23, 3 (1998), 438–458.

[62] Roy J Lewicki, Edward C Tomlinson, and Nicole Gillespie. 2006. Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of management* 32, 6 (2006), 991–1022.
[63] Michael Liebrenz, Roman Schleifer, Anna Buadze, Dinesh Bhugra, and Alexander Smith. 2023. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *The Lancet Digital Health* 5, 3 (2023), e105–e106.
[64] Zhuoran Lu and Ming Yin. 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*. 1–16.
[65] Joseph B Lyons, Katia Sycara, Michael Lewis, and August Capiola. 2021. Human–autonomy teaming: Definitions, debates, and directions. *Frontiers in Psychology* 12 (2021), 589585.
[66] Michael A Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.
[67] Stephen Marsh and Mark R Dibben. 2003. The role of trust in information science and technology. *Annual Review of Information Science and Technology (ARIST)* 37 (2003), 465–98.
[68] Joseph A Maxwell. 2012. *Qualitative research design: An interactive approach*. Sage publications.
[69] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.
[70] Roger C Mayer and Mark B Gavin. 2005. Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of management journal* 48, 5 (2005), 874–888.
[71] Daniel J McAllister. 1995. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal* 38, 1 (1995), 24–59.
[72] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 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 (2019), 1–23.
[73] Nathan McNeese, Mustafa Demir, Erin Chiou, Nancy Cooke, and Giovanni Yanikian. 2019. Understanding the role of trust in human-autonomy teaming. (2019).
[74] Nathan J. McNeese, Mustafa Demir, Erin K. Chiou, and Nancy J. Cooke. 2021. Trust and Team Performance in Human–Autonomy Teaming. *International Journal of Electronic Commerce* 25, 1 (Jan. 2021), 51–72. https://doi.org/10.1080/10864415.2021.1846854
[75] Nathan J McNeese, Mustafa Demir, Nancy J Cooke, and Christopher Myers. 2018. Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors* 60, 2 (2018), 262–273.
[76] Sharan B Merriam and Elizabeth J Tisdell. 2015. *Qualitative research: A guide to design and implementation*. John Wiley & Sons.
[77] Tim Merritt and Kevin McGee. 2012. Protecting artificial team-mates: more seems like less. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2793–2802.
[78] Debra Meyerson, Karl E Weick, Roderick M Kramer, et al. 1996. Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research* 166 (1996), 195.
[79] David R Millen. 2000. Rapid ethnography: time deepening strategies for HCI field research. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. 280–286.
[80] Dan Manh Nguyen. 2020. 1, 2, or 3 in a HAT? How a human-agent team's composition affects trust and cooperation. (2020).
[81] Kazuo Okamura and Seiji Yamada. 2020. Adaptive trust calibration for human-AI collaboration. *Plos one* 15, 2 (2020), e0229132.
[82] Thomas O'Neill, Nathan McNeese, Amy Barron, and Beau Schelble. 2022. Human–autonomy teaming: A review and analysis of the empirical literature. *Human factors* 64, 5 (2022), 904–938.
[83] Michael Pflanzer, Zachary Traylor, Joseph B Lyons, Veljko Dubljević, and Chang S Nam. 2023. Ethics in human–AI teaming: principles and perspectives. *AI and Ethics* 3, 3 (2023), 917–935.
[84] Minna Räsänen and James M Nyce. 2006. A new role for anthropology? rewriting" context" and" analysis" in HCI research. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. 175–184.
[85] Amy Rechkemmer and Ming Yin. 2022. When confidence meets accuracy: Exploring the effects of multiple performance indicators on trust in machine learning models. In *Proceedings of the 2022 chi conference on human factors in computing systems*. 1–14.
[86] John K Rempel, John G Holmes, and Mark P Zanna. 1985. Trust in close relationships. *Journal of personality and social psychology* 49, 1 (1985), 95.
[87] Tobias Rieger, Eileen Roesler, and Dietrich Manzey. 2022. Challenging presumed technological superiority when working with (artificial) colleagues. *Scientific Reports* 12, 1 (2022), 3768.
[88] Frank E Ritter, Nigel R Shadbolt, David Elliman, Richard M Young, Fernand Gobet, and Gordon D Baxter. 2003. Techniques for modeling human performance in synthetic environments: A supplementary review. *Human Systems*

Information Analysis Center, Wright-Patterson Air Force Base, Dayton, OH (2003).
[89] Julian B Rotter. 1980. Interpersonal trust, trustworthiness, and gullibility. *American psychologist* 35, 1 (1980), 1.
[90] Kristin E Schaefer, Edward R Straub, Jessie YC Chen, Joe Putney, and Arthur W Evans III. 2017. Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cognitive Systems Research* 46 (2017), 26–39.
[91] Paul Scharre. 2018. *Army of none: Autonomous weapons and the future of war*. WW Norton & Company.
[92] Beau G. Schelble, Christopher Flathmann, Nathan J. McNeese, Guo Freeman, and Rohit Mallick. 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 (Jan. 2022), 1–29. https://doi.org/10.1145/3492832
[93] Beau G Schelble, Jeremy Lopez, Claire Textor, Rui Zhang, Nathan J McNeese, Richard Pak, and Guo Freeman. 2022. Towards ethical AI: Empirically investigating dimensions of AI ethics, trust repair, and performance in human-AI teaming. *Human Factors* (2022), 00187208221116952.
[94] F David Schoorman, Roger C Mayer, and James H Davis. 2007. An integrative model of organizational trust: Past, present, and future. , 344–354 pages.
[95] Ben Shneiderman. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (2020), 495–504.
[96] Tony L. Simons and Randall S. Peterson. 1998. Task Conflict snd Relationship Conflict in Top Management Teams:The Pivotal Role of Intragroup Trust. https://api.semanticscholar.org/CorpusID:10732108
[97] Indramani L Singh, Robert Molloy, and Raja Parasuraman. 1993. Automation-induced" complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology* 3, 2 (1993), 111–122.
[98] Gabriel Szulanski, Rossella Cappetta, and Robert J Jensen. 2004. When and how trustworthiness matters: Knowledge transfer and the moderating effect of causal ambiguity. *Organization science* 15, 5 (2004), 600–613.
[99] Takane Ueno, Yuto Sawa, Yeongdae Kim, Jacqueline Urakami, Hiroki Oura, and Katie Seaborn. 2022. Trust in human-AI interaction: Scoping out models, measures, and methods. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–7.
[100] Anna-Sophie Ulfert. 2020. A Model of Team Trust in Human-Agent Teams. In *ICMI '20 Companion*. ACM, Virtual Event, Netherlands, 171–75. https://doi.org/10.1145/3395035.3425959
[101] Niels van Berkel and Vassilis Kostakos. 2021. Recommendations for conducting longitudinal experience sampling studies. *Advances in Longitudinal HCI Research* (2021), 59–78.
[102] James C Walliser, Ewart J de Visser, and Tyler H Shaw. 2016. Application of a system-wide trust strategy when supervising multiple autonomous agents. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 133–137.
[103] James C Walliser, Ewart J de Visser, Eva Wiese, and Tyler H Shaw. 2019. Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making* 13, 4 (2019), 258–278.
[104] Xinru Wang, Zhuoran Lu, and Ming Yin. 2022. Will you accept the ai recommendation? predicting human behavior in ai-assisted decision making. In *Proceedings of the ACM Web Conference 2022*. 1697–1708.
[105] Xinru Wang and Ming Yin. 2022. Effects of explanations in ai-assisted decision making: Principles and comparisons. *ACM Transactions on Interactive Intelligent Systems* 12, 4 (2022), 1–36.
[106] Magdalena Wischnewski, Nicole Krämer, and Emmanuel Müller. 2023. Measuring and Understanding Trust Calibrations for Automated Systems: A Survey of the State-Of-The-Art and Future Directions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
[107] Meng Xiao and Haibo Yi. 2021. Building an efficient artificial intelligence model for personalized training in colleges and universities. *Computer Applications in Engineering Education* 29, 2 (2021), 350–358.
[108] Kun Yu, Shlomo Berkovsky, Ronnie Taib, Jianlong Zhou, and Fang Chen. 2019. Do i trust my machine teammate? an investigation from perception to decision. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 460–468.
[109] Dale E Zand. 1972. Trust and managerial problem solving. *Administrative science quarterly* (1972), 229–239.
[110] Guanglu Zhang, Leah Chong, Kenneth Kotovsky, and Jonathan Cagan. 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 (Feb. 2023), 107536. https://doi.org/10.1016/j.chb.2022.107536
[111] Rui Zhang, Wen Duan, Nathan J. McNeese, Christopher Flathmann, Guo Freeman, and Alyssa Williams. 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 (2023), 1–31. https://doi.org/10.1145/3610072
[112] Rui Zhang, Christopher Flathmann, Geoff Musick, Beau Schelble, Nathan J McNeese, Bart Knijnenburg, and Wen Duan. 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 (2024), 1–23.

[113] Rui Zhang, Nathan J. McNeese, Guo Freeman, and Geoff Musick. 2021. "An Ideal Human": Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 1–25. https://doi.org/10.1145/3432945

[114] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 295–305.

[115] Roxanne Zolin, Pamela J Hinds, Renate Fruchter, and Raymond E Levitt. 2004. Interpersonal trust in cross-functional, geographically distributed work: A longitudinal study. *Information and organization* 14, 1 (2004), 1–26.

Received January 2024; revised April 2024; accepted May 2024