CHAPTER

13

Distributed dynamic team trust in human, artificial intelligence, and robot teaming

Lixiao Huang, Nancy J. Cooke, Robert S. Gutzwiller, Spring Berman, Erin K. Chiou, Mustafa Demir, and Wenlong Zhang

Center for Human, Artificial Intelligence, and Robot Teaming, Arizona State University, Tempe, AZ, United States

Introduction

A "Team" has been defined in the literature as an interdependent group of individuals, each with their own roles and responsibilities, who come together to address a specific goal (Salas, Dickinson, Converse, & Tannenbaum, 1992). When artificial agents (e.g., AI agents and embodied robotic agents) are added to an all-human team, the artificial agents may fulfill some roles and responsibilities, which are considered teaming. Therefore, reaching a high level of automation is not a defining feature for teaming capability. Instead, each agent requires varying levels of interdependency between it and other team members on a team task, and this interdependency defines teaming (Johnson & Vera, 2019).

Maintaining good interpersonal trust is important to any functional team. Lee and See's (2004) seminal trust model, for example, included cultural, organizational, and interpersonal factors in humans' trust in automation, but few studies have explored the interactions between these factors in depth. One such exception is Ho, Sadler, Hoffmann, Lyons, and Johnson's (2017) case study of the Automatic Ground Collision Avoidance System (Auto-GCAS). Auto-GCAS is a system that detects flight conditions and attempts to prevent aircraft from crashing into

https://doi.org/10.1016/B978-0-12-819472-0.00013-7

terrain, thereby saving pilots' lives. Different from most research on an end-user's trust in an artificial agent, Ho et al. (2017) provided rich descriptive information about real-world trust factors of three types of stakeholders: managers, engineers, and testing pilots. Managers were responsible for making decisions to implement the system in the military. If the system failed, they risked losing funding, equipment, and personnel. Engineers were responsible for developing the system to save pilots' lives when needed and otherwise not interrupt their tasks. If the system failed, engineers would risk strong reactions from other stakeholders, which might negatively impact engineers' careers and their emotions. The pilots were responsible for using the Auto-GCAS to save their own lives during operation, and as assistance toward completing missions successfully. The risk of failure for the pilot would be an unsuccessful mission and even death. Together, these separate risks and needs showcased how real-world automated systems impact a much broader, more diverse network of distributed stakeholders, beyond end-users. From this perspective, we believe that trust is also distributed across different stakeholders, in addition to dynamically changing, as each group member observes, uses, and communicates with one another about the system.

Existing trust research has focused on the end-users' perspective toward artificial agents like AI agents or embodied robotic agents, often neglecting other stakeholders' trust in the system and how that may impact the end-users' trust and use of the system. In Ho et al.'s (2017) example, the trustworthiness development of Auto-GCAS was an ongoing communication process among managers, engineers, and pilots. Engineers shared testing data with all stakeholders, and pilots expressed concerns to engineers. Managers could provide encouragement and training support for engineers and pilots. These communications among different groups could result in iterative improvements to system performance and the stakeholders' understanding of the system (a.k.a., mental models) and thus help all stakeholders develop appropriate trust in the system. The interrelated, communicative types of interpersonal factors in complex teams are not captured by trust studies that consider only individual end-users.

Therefore, we propose a framework, *Distributed Dynamic Team Trust (D2T2)*, to include both interpersonal factors and technical factors related to trust in human-AI-robot teaming along a dynamic timeline, all of which are typically missing in traditional dyadic trust research. The goal of this chapter is to invite more in-depth thinking and discussion about trust as a distributed, networked state that is constantly in flux, rather than a dyadic and static measure. The following sections will elaborate on D2T2, describe its potential measurement and modeling approach, and finally, discuss a set of applications that reflect its utility.

Distributed and dynamic team trust (D2T2)

Trust distributes among different types of stakeholders

As team sizes and compositions change, humans' trust in the artificial agents becomes distributed among all related stakeholders (see Fig. 1). Each stakeholder's attitude toward the artificial agent of interest can play a role in shaping team trust. Their trust in the Auto-GCAS may influence their job performance and, in turn, may influence team performance. The D2T2 framework allows for including other related stakeholders beyond the end-users, which is essential for providing a holistic view of trust development in the team (Ho et al., 2017).

Trust distributes within smaller teams of stakeholders

Different types of stakeholders may influence each other's trust in the artificial agents, and within each type of stakeholder, there are smaller networks using communication channels that influence one another's trust in the artificial agents (see Fig. 1). Communication about the AI agents is an important influencing mechanism, potentially more so among

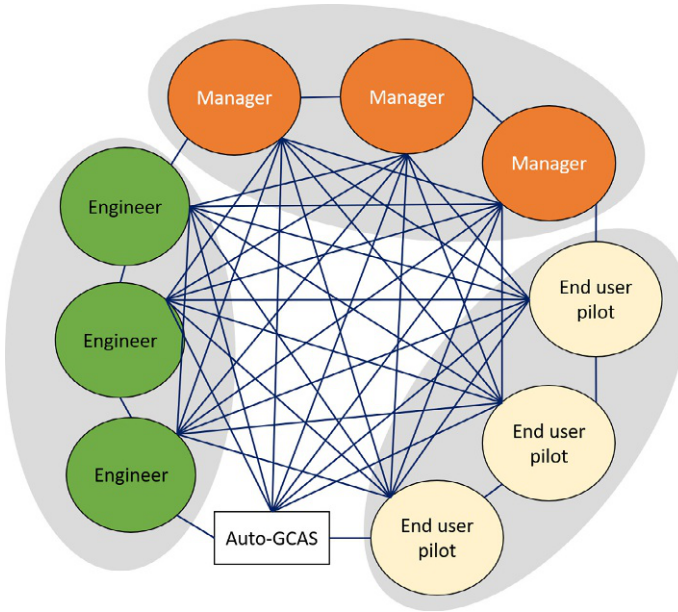

FIG. 1 Trust network in the Auto-GCAS Example. The colored nodes = stakeholders; the solid links = trust relationships; gray-shaded regions = smaller network.

stakeholders with similar roles. For example, in addition to the influence of engineers' shared testing data on managers' and pilots' trust in the Auto-GCAS system, word-of-mouth communications within the pilot network may also significantly influence pilots' trust in the Auto-GCAS system over time (Ho et al., 2017).

Distributed team trust is transitive, as well as dynamic

Interpersonal trust among all related stakeholders and their trust in artificial agents may be transmitted across the groups through daily conversations, newsletters and policies, and training procedures. The framework of D2T2 is an attempt to recognize and measure this potential impact on each type of stakeholders' trust in any artificial agent as well as one another. In an example later (Fig. 2), team trust can change through direct interactions with artificial agents or change indirectly through other people's influence.

Fig. 2 is a simplified illustration of trust transitivity in a distributed and dynamic team with two types of stakeholders—an end-user and a trainer, as well as an artificial agent. The end-user's initial relationship with the artificial agent is one of distrust (Phase 1). However, the end-user then interacts with a trainer, who has developed a positive trust relationship with the artificial agent; the end-user develops a positive trust relationship with this trainer (Phase 2). The trainer communicates new information about how and when to trust the artificial agent, and this, in turn, changes how the end-user views the artificial agent more positively (Phase 3).

Fig. 2 is a microcosm of the human-AI-robot system, which will certainly include more users, trainers, and other stakeholders. The importance of trust as a distributed and transitive concept arises under those conditions with more types of stakeholders and should be studied in addition to existing dyadic trust research. The potentially predictive relationships (in parentheses) for a variety of starting cases illustrate that end-users' trust in the artificial agent may change because of *relationships*

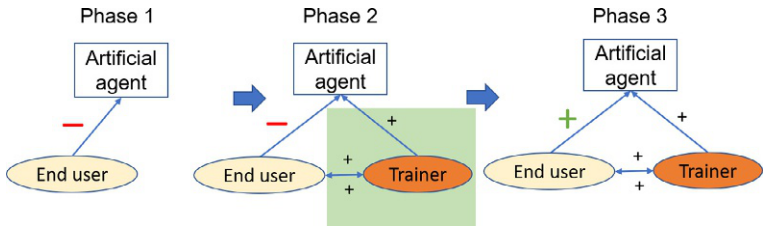

FIG. 2 Trust transitivity illustration. An end-user initially distrusts an artificial agent. The user begins interacting with a trainer. The trainer's relationship with the agent is positive, and because the end-user develops a positive trust relationship with the trainer, the end-user begins to trust the artificial agent.

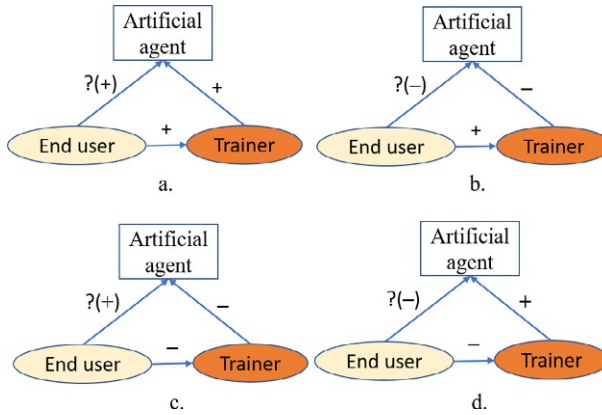

FIG. 3 Examples of inter-relational trust states. +/− = established trust status, arrow = trust direction, (+/−) = anticipated or uncertain result.

with other humans on the team (see Fig. 3). For example, in case (a), if the end-user trusts the trainer and the trainer trusts the agent, we might predict the end-user is more likely to trust the agent too. In case (b), the end-user trusts the trainer, but the trainer does not trust the agent, and we would predict the end-user may not trust the agent either, or the trainer's attitude toward the agent may reduce the end-user's trust in the agent. In examples (c) and (d), the end-user does not trust the trainer, so the trainer's trust in the agent may not play a role, and the transitivity may not exist (or it may be negative, in that the end-user decides to distrust everything in the system given that the end-user perceives the trainer as untrustworthy).

Trust transitivity has been observed in human-human trust (Jøsang, Hayward, & Pope, 2006), so it seems plausible that this transitivity applies to trust in a human-AI-robot team. The D2T2 framework takes this into account and suggests testable hypotheses about trust relationships in teams, proposing a new way to characterize and to measure team trust.

Because D2T2 involves a timeline and dynamics, we can also consider the stability of team trust, which refers to a state of trust across a distributed set of stakeholders that reaches and maintains a productive unified state over a given period. If trust within the network changes dramatically within the measurement period, it may indicate a problem. A D2T2-based trust network could indicate where and what that problem may be (e.g., decreased trust by one person for one artificial agent, in a network, would look very different than decreased trust by a group of people toward one artificial agent in an isolated setting). Viewing trust as a dynamic process that waxes and wanes over time, at times dropping rapidly, and in the context of the dynamic interactions among team members, may lead to more realistic, continuous metrics of trust.

The influence of stakeholders' views on others' trust may vary

In a distributed network, depending on the task context and people's roles and responsibilities, some individuals may have a bigger impact on others' attitudes than others. For instance, in a survival situation with high risk and uncertainty, a commander's calmness and order may greatly impact crew members' affection and beliefs (Torrance, 1954). Likewise, a senior trainer's negative experience and advice to avoid using an artificial agent in certain situations may override a more novice peer's positive perception of the artificial agent. Based on their trust in an artificial agent, managers' decision making may even determine whether an artificial agent is deployed for the end-users to use and build their trust. Therefore, the trust relationships of a few critical people in the network may influence whether the end-users have the opportunity to interact with the artificial agent and learn to use the artificial agent appropriately. For this reason, different stakeholders' degree of influence is also a variable to consider when analyzing D2T2, a nuance that is required when examining trust in teams.

D2T2 can cover multiple artificial agents on a team

In human-AI-robot teaming, it is likely that more than one artificial agent is developed or deployed on a team. Some artificial agents may be similar or even identical, while others may have totally different designs and functions. In the context of an end-user working with a swarm of homogeneous robots that are identical, the end-user's trust in all the robots is expected to be the same. However, a study found soldiers develop an emotional attachment to some robots based on long-term interactions on the battlefield, and especially when the robot was destroyed in place of the soldier during a mission (Carpenter, 2013). Anecdotal evidence suggests that other identical replacement robots may not gain the same level of affection (Carpenter, 2013). Because affection plays a critical role in forming trust (Lee & See, 2004), the other identical robots may not gain the same level of trust as the sacrificed one. This indicates the necessity to study distributed trust in the multiple artificial agents that stakeholders interact with on a team, not just one artificial agent. It also demonstrates that a more dynamic perspective on trust is necessary, one that would capture the experiences of the stakeholder with a specific robot, as well as the differences between the trusted robot and other less trusted but mechanically identical robots, to evaluate how trust may evolve differently with these robots over time.

Distributed dynamic team trust also applies to different artificial agents people interact with. The differences between different artificial agents may include versions of the agents due to software upgrades or added

features. For example, an automatic target recognition (ATR) technique with 70% reliability may be acceptable to soldiers in an early stage but may not be as satisfying as 95% reliability when it becomes possible. Knowing that a 95% reliability option is available or having interacted with such a high-reliability ATR may reduce soldiers' trust in the original lower reliability ATR, such that it affects operations. Therefore, a research question for D2T2 is how the knowledge of, or order of experiencing artificial agents with different levels of capabilities can change different types of stakeholders' trust in the artificial agent in networked groups.

Because system versions and upgrades do not always happen in parallel across military platforms, warfighters who are deployed across different variants of an artificial agent could struggle to develop calibrated trust. Meanwhile, when a team involves multiple heterogenous artificial agents, whether and how trust in one type of artificial agent transfers to another type is a question related to trust transfer in D2T2 and may depend on the similarities and differences between the artificial agents. In the case of Auto-GCAS, a manager's and a pilot's negative predispositional trust in the GCAS came from their previous experience with a separate unreliable system (Ho et al., 2017). It is not clear how different the separate unreliable system was from the Auto-GCAS system and how the stakeholders' trust in a different system was transferred across. It suggests trust could migrate across experiences with different types of systems within a domain as well, and not just upgrades made to existing systems. It is unlikely that static or dyadic measures of trust can capture these influences well. The D2T2 framework may allow this, and we expand on potential methods for measuring and modeling this framework of team trust next.

Measurement and modeling of distributed dynamic team trust

The approach to addressing the D2T2 framework in human-AI-robot teaming starts with understanding the context, including the tasks, environment, the stakeholders, and artificial agents involved. We must also understand the kinds of interactions that are available between the entities involved in a situated context, where transitive properties of trust take place. These affordances for interaction are also where we begin to measure trust in a distributed and dynamic fashion, through the different measures that may be appropriate for different types of trust relationships. We outline the steps later using the context of Auto-GCAS (Ho et al., 2017) as an example.

Step 1: Identify the context of interest. The context of interest refers to the position domain, such as a specific position in an application field. For example, the US Air Force was interested in flight mission success and

pilots' safety, so a military flight crew and their safety in flying was the context of interest. In many incidents of controlled flights into terrain (CFIT), the crashes of healthy and functional aircraft are due to pilot errors, physical challenges, and mental challenges (Ho et al., 2017). Therefore, a goal was to develop an artificial agent to aid the pilots and prevent CFIT. Identifying this context links trust to domain expertise and existing knowledge about the people, capabilities of the artificial agent, and goals of the task environment. Naturally, this plays a large role in developing trustworthy technology but also serves to base trust assessment in realistic, grounded methodology.

Step 2: Identify stakeholders and the artificial agents in the trust network and how they interact with one another. The Auto-GCAS example (Ho et al., 2017) identified one artificial agent and three types of stakeholders. Then we can interview subject matter experts and review training documents to do task analysis to identify related stakeholders, technologies, and interactions in a situated context, such as rail and air dispatching (Huang, Cummings, & Nneji, 2018; Huang, Nneji, & Cummings, 2019). It is necessary to understand these interactions in more than high-level detail to make more fine-grained predictions or assessments. More nuanced methods and elicitation techniques have been useful as the foundation for characterizing teams as networks, such as Event Analysis of Systemic Teamwork (EAST; Walker et al., 2006).

Step 3: Measure each trust relationship in the trust network. A challenging problem in realizing D2T2 is how to characterize so many different trust relationships in a network and how to measure them with enough frequency over time. Our strategy is to investigate D2T2 by system stages, layers of trust, and the types of interactions.

System stages refer to qualitatively different event cutoff lines in a work domain. System stages determine the interactions among the entities. For example, the findings in Ho et al. (2017) were based on the predeployment stage alone, which did not involve operating the fully functioning Auto-GCAS system yet. The predeployment stage would naturally be different from the training stage when pilots use Auto-GCAS in a simulation or the operation stage when pilots operate a real aircraft. To mitigate the complexity of the analysis, we propose to examine the trust relationships, including trust transitivity and the trust network of stakeholders, by system stages. These stages can then be linked or compared (see Step 4) when more data about relevant trust relationships among stakeholders and artificial agents become available.

Within each system stage, individuals may have three layers of trust: predispositional trust—before operational interactions, learned trust—trust during a controlled training setting or based on ones' reflection of past experiences with the artificial agent, and situational trust—trust during real-life interactions with an artificial agent when unpredictable things

may happen (Hoff & Bashir, 2015). Therefore, the D2T2 approach should also track the layer of trust in a given relationship.

Moreover, just as important is that each layer may need a different trust measure and may require different administrations, depending on what type of interactions are involved. Our hypothesis is that interaction affordances in human-human relationships and in human-AI-robot relationships may be different, and thus some trust measures may be more suitable than others in each trust relationship. For example, text-based analysis of conversations may be one way to unobtrusively measure trust between people (Lee & Kolodge, 2019), but interacting with an artificial agent that does not have natural language abilities may preclude the use of such measures. Likewise, surveys are good for predispositional trust, whereas unobtrusive measures like behavioral indicators may be more feasible than surveys for situational trust during ongoing activities. The next section on trust measures elaborates selected trust measurement options and comments on their applicability. However, current methods of trust measurement will have to be improved or adjusted to achieve a more distributed, dynamic assessment of team trust.

Step 4: Methods for assessing and analyzing a trust network. After measuring stakeholder's trust in each artificial agent using appropriate measures, we can gain a better understanding of team trust and its evolution using social network analysis methods to examine the trust network. The distributed team trust network may be compared by system stages, as discussed in Step 3. The fully connected trust network in Fig. 1 may also be customized to show critical trust relationships depending on the roles of stakeholders because not all influencing factors for trust equally matter to all stakeholders and thus can be filtered case by case (Ho et al., 2017). Team trust networks can show different properties of trust, such as the distribution of team trust, trust transitivity, trust transfer, changing patterns of the trust network, power or influences of some stakeholders, and dynamic effects. Each of these has importance in a better understanding of trust and teaming with a given artificial agent or several agents. Most importantly, these kinds of assessments may guide a strategy for trust repair or recalibration—changes in training, direction from supervisors, the information delivered to end-users, transparent displays, and other options all present solutions, but the assessments and strategies should match the suspected problem and the roles of stakeholders.

Trust measures

As we discussed in Step 3, the framework of D2T2 brings the challenge of dynamic and distributed trust measurement techniques. These are required to implement the framework and provide value. Dynamic team

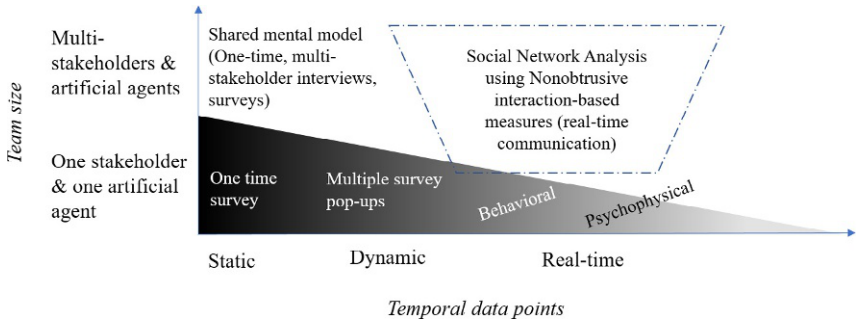

FIG. 4 Trust measurement needs for D2T2. The X-axis indicates the relative quantity of temporal data points that are required for each general type of measure. The Y-axis shows the size of teams directly involved in the trust measurement. Most existing research focuses on one human's trust in one artificial agent, as shown in the gray area; darkness (and thickness) indicate the number of studies, the darker (and the thicker), the more. The dotted box indicates interactive cognition theory related methods, which may apply to both dyadic and team trust.

trust means team trust changes over time and should be measured and modeled accordingly. The dynamic time scale could be seconds, minutes, hours, days, or even in terms of missions and events, depending on the task of interest, and the readiness of data collection techniques. In considering the measurement options for the trust links within the framework of D2T2, two dimensions emerge: (a) the number of entities included in a measure of trust, and (b) the number of temporal data points collected by it (see Fig. 4). We introduce each measure in more detail later.

Trust surveys

Surveys are the most common measure of human trust. These self-reported surveys are typically administered one time, either before or following an intervention, depending on whether the goal is to measure baseline trust or trust in a specific technology. The surveys often involve the use of Likert scales and ordinal data (e.g., Jian, Bisantz, & Drury, 2000), with many adaptations available throughout the literature (Hancock et al., 2011; Hoff & Bashir, 2015). Surveys focus on one person's perceptions of another entity, including trust in automation across different domains (Pak, Rovira, McLaughlin, & Baldwin, 2017), interpersonal trust (Rotter, 1967), and trust in specific contexts, such as small military teams (Adams, Bruyn, & Chung-yan, 2004). Surveys will not fully realize the D2T2 framework, in terms of the dynamic process of changes in trust, as they are often prohibitive to re-administer at a high rate and will not capture communications between agents. Surveys will be poor at feeding into any real-time assessments and will be interruptive to the performance of operators and others in the network. Even when surveys are useful, all

measures must contend with validity and reliability of measurement concerns (Gutzwiller et al., 2019; Jian et al., 2000; Spain, Bustamante, & Bliss, 2008). However, surveys are the most tested trust measure, and maybe most appropriate for measuring dispositional trust—relatively stable attitudes toward an agent—in a slow-paced, low-stress environment. Surveys may also measure learned trust immediately following a training intervention about how to work with an artificial agent. Therefore, surveys could be incorporated and may be useful in assessing distribution of trust and its transitive properties.

Binary behavioral indicators of trust

Existing research has used various behavioral measures to infer trust. A common and simplified approach is using binary use and disuse as an indicator of trust. In a study using a simulated robotic command and control system, the number of human overrides of robotic behavior reflected an operator's trust in their robot partners (Freedy, DeVisser, Weltman, & Coeyman, 2007). However, behavioral indicators of reliance, compliance, and intervening actions (Domeyer, Venkatraman, Price, & Lee, 2018) may be observable in both the absence or presence of trust. If human operators have the option not to use the artificial agent, they may avoid using the agent when they distrust it; but if the use of the agent is required to do the task, they may still use it even though they do not trust the agent.

At that level of analysis, it may generate more temporal data than surveys but does not reflect socially dynamic and distributed trust at a team level. However, as not all members will be directly interacting, the measure is limited in larger networks of distributed agents. In addition, an issue with binary behavioral measures of trust is that some behavioral outcomes occur infrequently or may be too delayed to analyze. For example, binary behavioral indicators of trust (e.g., authorization versus nonauthorization of actions like missile launches) cannot occur early enough to prevent consequences of inappropriate or degraded trust. Therefore, this measure has limited usability to realize D2T2 but may remain useful for dyadic cases in which users have a choice not to use the artificial agent, their use or disuse happens infrequently, and the action of use or disuse would not lead to immediate irrevocable catastrophe.

Sensor-based psychophysical measures of trust

According to the comparison of trust measures (see Fig. 4), real-time sensor-based psychophysical measures of trust have the potential of collecting continuous data, providing more granularity to the dynamic aspects of trust. Developers of artificial agents seek this kind of data to develop algorithms that enable artificial agents to detect and adjust responses to the changes in humans' trust; a discrete and static measure

of trust through surveys cannot satisfy this purpose. Recent work has proposed to use electroencephalography (EEG) and Galvanic skin response (GSR), but the relationship between psychophysical data and trust has not been well established yet.

Electroencephalography has been a popular physical measure for trust estimation (Boudreau, McCubbins, & Coulson, 2009; Kolling, Walker, Chakraborty, Sycara, & Lewis, 2016; Long, Jiang, & Zhou, 2012; Mondada, Karim, & Mondada, 2016). Human trust had a significantly positive correlation with EEG features from specific electrodes in an investment game with artificial agents, and the frontal and occipital areas were identified as the predominant brain areas correlated with trust (M. Wang, Hussein, Rojas, Shafi, & Abbass, 2018).

Galvanic skin response captures physiological arousal levels based on the conductivity of the surface of the skin (Akash, Hu, Jain, & Reid, 2018). GSR is significantly affected by both trust and cognitive load in a text-chat environment (Khawaji, Zhou, Chen, & Marcus, 2015), and the phasic component of GSR was also a significant predictor of trust levels (Akash et al., 2018), but the results were not clear about the corresponding GSR readings for a high level of trust.

In general, sensor-based measurement of human trust is still in its infancy, not well explored and validated to use such real-time estimates of human trust. One issue is that most of the sensors used for measuring trust require careful calibration, and almost all the results are collected using interactions with software-based agents. It remains an open question whether such sensing methods can provide reliable results for interactions between humans and physical robots in outdoor and complex environments (Wang et al., 2019). Other issues arise from the challenges associated with the sensors themselves, including sensing noise and high computation load. Under the D2T2 framework, we must discover how to bind these measures to capture the trust of the interaction between a human and any *specific* agent. For example, it may be difficult to identify trust measurement from sensor-based signals separately when the participant has multiple tasks and interacts with multiple agents during data collection. More evidence is needed to show the validity and reliability of using sensor-based data in measuring trust to facilitate its use in the D2T2 measurement solution.

In sum, surveys, binary behavioral indicators, and psychophysical measures of trust focus on trust in a dyadic relationship. These measures are also somewhat context free, which makes sense given their originating purpose (e.g., a survey given after interaction with a system in a lab). Understanding the interaction processes and team cognition and moving beyond individual-centered orientations are essential for effective human-AI-robot teaming (Johnson & Vera, 2019). To this end, we introduce interaction-based measures of trust.

Interactive team cognition theory and interaction-based measures

Team cognition has been conceived of in terms of shared mental models (Cannon-Bowers, Salas, & Converse, 1993), distributed cognition (Hutchins, 1995), or transactive memory (Hollingshead, 1998). Though many interesting findings, theoretical developments, and measures have resulted from these perspectives, they each take a static view of team cognition, measuring it as a set of static snapshots in time. Adopting the shared mental model perspective for team trust would suggest that team trust is the interconnected pairwise trust of team members at a specific moment. In a shared mental model (SMM) framework, each human teammate's trust would be measured for each dyad and then aggregated to reflect a *full* picture of team trust in a distributed social network. A trust network generated this way takes a holistic view of team trust at a specific time, but it critically misses the dynamic aspects of transitive trust mentioned earlier.

A more dynamic way of thinking about team cognition is through the theory of interactive team cognition (ITC; Cooke, 2015; Cooke, Gorman, Myers, & Duran, 2013). ITC posits that team cognition is an activity that consists of interactions among teammates, it ties strongly to the task context and the modality of interactions, and it is the best measured at the team level of analysis. Thinking about team cognition this way has resulted in new and interesting findings, theoretical developments, and measures (Cooke & Gorman, 2009; Grimm, Demir, Gorman, and Cooke, 2018; Gorman, Demir, Cooke, & Grimm, 2019; McNeese, Demir, Cooke, & Myers, 2018). Team interactions may include continuous behavioral interaction data and verbal communication. In particular, new process-oriented measures of team cognition have been developed by relying heavily on the analysis of communication dynamics. Communication data are rich and can be examined over multiple dimensions, such as volume, frequency, pitch, content, speech act, and flow or who is talking to whom (Cooke & Gorman, 2009). Communication analysis in experimentation is unobtrusive and possible to collect and analyze in real time automatically.

In addition, unlike static snapshots, communication dynamics can reveal changes in a team after a perturbation, including periods of adaptation and resilience (Cooke & Gorman, 2009). Communication may include real-time vocal conversation, text chat messages, social media posts, and newsletters. To measure trust this way, cognitive indicators of trust that could be found in team interactions, such as communications, should be identified and measured over time. For example, certain communication content and flow may indicate a lack of trust. Researchers also found a positive relationship between the number of social messages exchanged in virtual, temporary teams, and swift trust (Jarvenpaa & Leidner, 1999). In team trust, communication may influence trust calibration by encouraging learning about the artificial agents' performance or

other properties in several ways (and recall our earlier discussion of transitive trust, which could occur through human-human communications among managers, engineers, and pilots).

Using social network analysis to model D2T2

Ho et al.'s (2017) case study provided a qualitative analysis example to explore how interpersonal relationships may influence different stakeholders' trust in the Auto-GCAS through various communication channels. As shown in Fig. 1, a social network contains links and nodes. Each node represents a stakeholder or an artificial agent. Each link in the fully connected trust network represents a trust relationship between any two entities, such as managers, engineers, pilots, and the Auto-GCAS system. Each trust relationship may have a suitable way to measure, depending on the system stage, the layer of trust, and the interactions between the entities.

With the social network analysis approach, we can study multiple trust relationships simultaneously. For example, communication analysis covers all related stakeholders and artificial agents, communication direction and frequency, and valence of the contents (i.e., positive or negative). The social network may show these trust relationships. Automating social network analysis is possible, but it requires a big data set of communication records and a fixed number of stakeholders to generate stable communication patterns on trust or distrust. In our current state, exploratory methods (e.g., interviews, surveys, social media posts, and work-related forum posts) are probably still necessary initially to understand the context and to identify critical indicators in communication for trust.

All in all, the selected trust measures listed in this chapter do not exclude other unobtrusive real-time measures for team trust dynamics. Further research is invited to identify dynamic patterns or indicators of changes in trust using the framework of D2T2.

Applications of distributed dynamic team trust in human-AI-robot teaming

The holistic view of D2T2 has broad applications in trust research because human-AI-robot teams are likely to comprise different types of stakeholders and multiple artificial agents, requiring an extension of the traditional examination of trust as a dyadic construct. The measurement and modeling approach still needs exploration and validation. Our major objective is to point out what is currently missing in prior measures and what our new approach provides to fill that gap. Below are several potential use cases that could benefit from the application of the D2T2 framework and conceptualization of team trust.

Identifying problematic artificial agents in a human-AI-robot team trust network

An artificial agent's trustworthiness is one property of the agent. Working with an untrustworthy agent may lead to poor team performance. In a heterogeneous team that includes multiple artificial agents, measuring an evidence-based and weighted trust in each agent from multiple stakeholders may help identify good or bad agents. This is similar to the use of multisource feedback and appraisal in organizational psychology (Fletcher, 2001), in which supervisors, peers, and customers all provide ratings to evaluate an employee's performance. Ho et al. (2017) found that the reputation of Auto-GCAS was developed through sharing engineers' testing data, managers' discussions, and pilots' word-of-mouth communications, which are all suitable candidates for measures of trust in our framework.

Further, incident reports, communication about the usage of the artificial agents, and system operation logs, which often already exist, could be utilized to evaluate the performance and acceptance of each artificial agent. A problematic agent could be fixed or eliminated from the team. Fixes could take a variety of forms, including appropriately targeted training of those stakeholders who are affecting the network most significantly or closest to the operational context. Such training could communicate the issue and ground the team in the same mental model or correct or educate those who do not yet view the agent accurately.

Using only a dyadic trust approach, some problematic agents may be nearly impossible to identify. However, when viewing team trust as a distributed state, it becomes possible to identify whether even one human who distrusts an agent may be infecting the entire team (or could be the lone dissenter). A D2T2 approach, as a more dynamic method, could help unearth these types of relationships and their dynamics more frequently and with sufficient detail to devise solutions. The use of more continuous, dynamic measures of trust (e.g., continual behavioral indicators or interactive communication analysis) may enable the unobtrusive detection of issues in team trust and the notification of team leaders or other stakeholders, serving as a new capability via diagnostic aid for those looking to make teams more effective.

Design of scalable control strategies for multiagent systems

A second potential application is the development of control frameworks for large-scale multiagent systems, such as robotic swarms. These control frameworks should rely on limited human supervision, in part because situation awareness and task performance can be degraded when humans team with large numbers of robotic agents at different levels of

automation (Chen & Barnes, 2014). Human supervision can be exercised through high-level specifications for the multiagent system, which allows the control approach to scale well with the number of agents (Michael, Fink, Loizou, & Kumar, 2010). From the perspective of fostering appropriate trust in the system, one important question is the optimal integration of such human-issued directives with individual and collective decisions by the robotic agents.

In the D2T2 framework, stakeholders need to know the similarities and differences among the artificial agents in the swarm, instead of the properties of one artificial agent. This knowledge may help end-users and other stakeholders to establish their trust criteria. For example, end-users may build trust in unused robots based on the robots they have already interacted with if all robots in the swarm have similar capabilities. Constructing time-varying networks of trust among stakeholders and multiple artificial agents can also aid in illustrating the relationships among teammates and bringing problematic ones to notice. For instance, control strategies for robotic swarms will require designating team leaders who supervise the swarm, and it would not be effective to assign team members with inappropriate trust in some of the robots to supervise them. Moreover, identifying key leaders of a large team in terms of which teammates trust them may help ascertain which teammates to support or request assistance from when determining a plan of action. For example, the ability to identify *influencers* of the swarm or of specific agents in the swarm may be useful to target those roles for trust training, intervention, etc., rather than being organizationally forced to train everyone.

Conclusion

Team trust is not an isolated phenomenon nor an outcome, but rather an interactive social and dynamic process that has multiple sources of influence over time (Hoff & Bashir, 2015; Lee & See, 2004). Current individual, end-user-oriented trust research neglects the different types of stakeholders' trust in one or more artificial agents, and how one stakeholder's trust in an artificial agent influences other stakeholders' trust, and vice versa. We proposed using the framework of distributed and dynamic team trust (D2T2) to fill the gaps in the existing trust research. We also proposed four steps to implement D2T2 in a human-AI-robot teaming context, including utilizing appropriate traditional trust measures for different layers of individual end-user trust and proposing interaction-based unobtrusive measures for capturing the distributed and dynamic property of team trust.

D2T2 contributes to the trust research literature by filling the gap of measuring and modeling the *distributed* interrelational trust between

human teammates and between human and artificial agents over time. We suspect that the D2T2 framework and modeling approach may move the field much closer to pragmatic and real-world trust measurements for complex systems and more advanced human-AI-robot teaming. The inclusion of team communications, distribution of trust beliefs, and network analysis enables much more appropriate measures of trust across large teams than current trust measures. However, we are aware that D2T2 is an exploratory framework, and await further experimentation, application, and development before suggesting any grand conclusions. As with all studies of trust in technology, it is reasonable to suggest that D2T2 in human-AI-robot teaming is vital to team performance—with the potential to proactively and potentially in real-time help to identify problematic artificial agents and call for supporting actions. If continuous and real-time trust measures are improved, this identification may even be done automatically in certain contexts. Lastly, empirical, interdisciplinary research is needed to verify the value of D2T2 when applied to real-world team situations and to identify novel ways of measuring and modeling D2T2.

References

Adams, B., Bruyn, L., & Chung-yan, G. (2004). *Creating a measure of trust in small military teams.* DRDC Toronto Technical Report CR 2004-077.

Akash, K., Hu, W.-L., Jain, N., & Reid, T. (2018). A classification model for sensing human trust in machines using EEG and GSR. *ACM Transactions on Interactive Intelligent Systems*, *8*(4), 1–20. https://doi.org/10.1145/3132743.

Boudreau, C., McCubbins, M. D., & Coulson, S. (2009). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Social Cognitive and Affective Neuroscience*, *4*(1), 23–34. https://doi.org/10.1093/scan/nsn034.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In *Current issues in individual and group decision making* (pp. 221–246). Hillsdale, NJ: Lawrence Erlbaum.

Carpenter, J. (2013). *The quiet professional: An investigation of US military explosive ordnance disposal personnel interactions with everyday field robots. (2013).* https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/24197/Carpenter_washington_0250E_12154.pdf?sequence=1.

Chen, J. Y., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13–29.

Cooke, N. J. (2015). Team cognition as interaction. *Current Directions in Psychological Science*, *24*(6), 415–419.

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Journal of Cognitive Engineering and Decision Making*, *3*(1), 27–46.

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

Domeyer, J., Venkatraman, V., Price, M., & Lee, J. D. (2018). Characterizing driver trust in vehicle control algorithm parameters. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*, 1821–1825.

Fletcher, C. (2001). Performance appraisal and management: The developing research agenda. *Journal of Occupational and Organizational Psychology*, *74*(4), 473–487.

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. In *2007 International symposium on collaborative technologies and systems* (pp. 106–114).

Gorman, J. C., Demir, M., Cooke, N. J., & Grimm, D. A. (2019). Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: A layered dynamics approach. *Ergonomics*, *62*(5), 629–643.

Grimm, D., Demir, M., Gorman, J. C., & Cooke, N. J. (2018). *The complex dynamics of team situation awareness in human-autonomy teaming. In 2018* IEEE conference on cognitive and computational aspects of situation management (CogSIMA) (pp. 103–109). (2018). https://doi.org/10.1109/COGSIMA.2018.8423990.

Gutzwiller, R. S., Chiou, E. K., Craig, S. D., Lewis, C. M., Lematta, G. J., & Hsiung, C.-P. (2019). Positive bias in the 'Trust in Automated Systems Survey'? An examination of the Jian et al. (2000) scale. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*, 217–221.

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

Ho, N. T., Sadler, G. G., Hoffmann, L. C., Lyons, J. B., & Johnson, W. W. (2017). Trust of a military automated system in an operational context. *Military Psychology*, *29*, 524–542.

Hoff, K., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434.

Hollingshead, A. B. (1998). Retrieval processes in transactive memory systems. *Journal of Personality and Social Psychology*, *74*(3), 659–671. https://doi.org/10.1037/0022-3514.74.3.659.

Huang, L., Cummings, M., & Nneji, V. C. (2018). Preliminary analysis and simulation of railroad dispatcher workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*, 691–695.

Huang, L., Nneji, V., & Cummings, M. (2019). How airline dispatchers manage flights: A task analysis in distributed and heterogeneous network operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 1389–1393. (2019). https://journals.sagepub.com/doi/10.1177/1071181319631182#articleCitationDownloadContainer.

Hutchins, E. (1995). *Cognition in the wild.* MIT Press.

Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science*, *10*(6), 791–815. https://doi.org/10.1287/orsc.10.6.791.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71.

Johnson, M., & Vera, A. (2019). No AI is an island: The case for teaming intelligence. *AI Magazine*, *40*(1), 16–28.

Jøsang, A., Hayward, R., & Pope, S. (2006). Trust network analysis with subjective logic. *Proceedings of the 29th Australasian Computer Science Conference*, *48*, 85–94.

Khawaji, A., Zhou, J., Chen, F., & Marcus, N. (2015). Using galvanic skin response (GSR) to measure trust and cognitive load in the text-chat environment. In: *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems, 1989–1994*. https://doi.org/10.1145/2702613.2732766.

Kolling, A., Walker, P., Chakraborty, N., Sycara, K., & Lewis, M. (2016). Human interaction with robot swarms: A survey. *IEEE Transactions on Human-Machine Systems*, *46*(1), 9–26.

Lee, J. D., & Kolodge, K. (2019). Exploring trust in self-driving vehicles through text analysis. *Human Factors*, 0018720819872672.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392.

Long, Y., Jiang, X., & Zhou, X. (2012). To believe or not to believe: Trust choice modulates brain responses in outcome evaluation. *Neuroscience*, *200*, 50–58. https://doi.org/10.1016/j.neuroscience.2011.10.035.

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

Michael, N., Fink, J., Loizou, S., & Kumar, V. (2010). Architecture, abstractions, and algorithms for controlling large teams of robots: Experimental testbed and results. In *Robotics research* (pp. 409–419). Springer.

Mondada, L., Karim, M. E., & Mondada, F. (2016). Electroencephalography as implicit communication channel for proximal interaction between humans and robot swarms. *Swarm Intelligence*, *10*(4), 247–265.

Pak, R., Rovira, E., McLaughlin, A. C., & Baldwin, N. (2017). Does the domain of technology impact user trust? Investigating trust in automation across different consumer-oriented domains in young adults, military, and older adults. *Theoretical Issues in Ergonomics Science*, *18*(3), 199–220. https://doi.org/10.1080/1463922X.2016.1175523.

Rotter, J. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality*, *35*(4), 651–665.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey, & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Ablex Publishing.

Spain, R. D., Bustamante, E. A., & Bliss, J. P. (2008). Towards an empirically developed scale for system trust: Take two. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52*, 1335–1339.

Torrance, E. P. (1954). The behavior of small groups under the stress conditions of "survival". *American Sociological Review*, *19*(6), 751–755.

Walker, G. H., Gibson, H., Stanton, N., Baber, C., Salmon, P., & Green, D. (2006). Event analysis of systemic teamwork (EAST): A novel integration of ergonomics methods to analyse C4i activity. *Ergonomics*, *49*(12–13), 1345–1369.

Wang, M., Hussein, A., Rojas, R. F., Shafi, K., & Abbass, H. A. (2018). EEG-based neural correlates of trust in human-autonomy interaction. In: *2018 IEEE symposium series on computational intelligence (SSCI)*, pp. 350–357. https://doi.org/10.1109/SSCI.2018.8628649.

Wang, Y., Lematta, G. J., Hsiung, C.-P., Rahm, K. A., Chiou, E. K., & Zhang, W. (2019). Quantitative modeling and analysis of reliance in physical human–machine coordination. *Journal of Mechanisms and Robotics*. *11*(6). https://doi.org/10.1115/1.4044545.