



ARL-TR-9461 • MAY 2022



Trust Measurement in Human–Autonomy Teams: Development of a Conceptual Toolkit

**by Andrea Krausman, Catherine Neubauer, Daniel Forster,
Shan Lakhmani, Anthony L Baker, Sean M Fitzhugh, Gregory
Gremillion, Julia L Wright, Jason S Metcalfe, and Kristin E
Schaefer**

Approved for public release: distribution unlimited.

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.



Trust Measurement in Human–Autonomy Teams: Development of a Conceptual Toolkit

**Andrea Krausman, Catherine Neubauer, Daniel Forster, Shan
Lakhmani, Sean M Fitzhugh, Gregory Gremillion, Julia L Wright, Jason
S Metcalfe, and Kristin E Schaefer**
DEVCOM Army Research Laboratory

Anthony L Baker
Oak Ridge Associated Universities; DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) May 2022		2. REPORT TYPE Technical Report		3. DATES COVERED (From - To) September 2020–March 2022	
4. TITLE AND SUBTITLE Trust Measurement in Human–Autonomy Teams: Development of a Conceptual Toolkit				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Andrea Krausman, Catherine Neubauer, Daniel Forster, Shan Lakhmani, Anthony L Baker, Sean M Fitzhugh, Gregory Gremillion, Julia L Wright, Jason S Metcalfe, and Kristin E Schaefer				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) DEVCOM Army Research Laboratory ATTN: FCDD-RLH-FD Aberdeen Proving Ground, MD 21005				8. PERFORMING ORGANIZATION REPORT NUMBER ARL-TR-9461	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release: distribution unlimited.					
13. SUPPLEMENTARY NOTES ORCID IDs: Andrea Krausman, 0000-0003-1955-8867; Sean Fitzhugh, 0000-0002-6283-2895; Daniel Forster, 0000-0001-8351-2009; Gregory Gremillion, 0000-0002-0205-688X; Shan Lakhmani, 0000-0001-6052-439X; Jason Metcalfe, 0000-0001-9086-9962; Catherine Neubauer, 0000-0002-6686-3576; Kristin Schaefer, 0000-0002-1342-3446; Julia Wright, 0000-0003-3026-1538					
14. ABSTRACT The rise in artificial intelligence capabilities in autonomy-enabled systems and robotics has pushed research to address the unique nature of human–autonomy team collaboration. The goal of these advanced technologies is to enable rapid decision making, enhance situation awareness, promote shared understanding, and improve team dynamics. Simultaneously, use of these technologies is expected to reduce risk to those who collaborate with these systems. Yet, for appropriate human–autonomy teaming to take place, especially as we move beyond dyadic partnerships, proper calibration of team trust is needed to effectively coordinate interactions during high-risk operations. But to meet this end, critical measures of team trust for this new dynamic of human–autonomy teams are needed. This report seeks to expand on trust measurement principles and the foundation of human–autonomy teaming to propose a “toolkit” of novel methods that support the development, maintenance, and calibration of trust in human–autonomy teams operating within uncertain, risky, and dynamic environments.					
15. SUBJECT TERMS human–computer interaction, HCI theory, concepts, and models, cross-computing tools and techniques, cross-computing tools and techniques, measurement, human–autonomy teaming, team trust, trust measurement, robot					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 103	19a. NAME OF RESPONSIBLE PERSON Andrea Krausman
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (Include area code) (410) 278-5933

Contents

List of Figures	v
List of Tables	vi
1. Introduction	1
2. What Is a Human–Autonomy Team?	2
2.1 Team Trust (Human–Autonomy Teams)	4
2.1.1 Implications of Interpersonal Trust for Human–Autonomy Teams	4
2.1.2 Implications of Human–Automation Trust for Human–Autonomy Teams	5
2.1.3 Implications of Information Trust for Human–Autonomy Teams	7
3. Team Trust Measurement for Human–Autonomy Teams	8
3.1 Types of Trust and Implications for Measurement	8
3.1.1 Trust Propensity	10
3.1.2 Affective Trust	10
3.1.3 Perceived Trustworthiness	11
3.1.4 Cognitive-based Trust	12
3.1.5 Situational Trust	12
3.1.6 Learned Trust	13
3.2 Trust Propagation and its Implications for Measurement	14
4. Human–Autonomy Team Trust Measurement Toolkit	15
4.1 Self-Report Measurement	15
4.1.1 Trust Propensity Scales	15
4.1.2 Trust in Autonomy	19
4.2 Behavioral Indicators	25
4.2.1 Attentional Monitoring	26
4.2.2 Reliance and Compliance	27
4.2.3 Proximity	29

4.2.4	Entrainment: Behavioral Synchrony	31
4.2.5	Facial Analysis	34
4.3	Communication Measures	36
4.3.1	Communication Rates	36
4.3.2	Communication Flow	37
4.3.3	Communication Content	39
4.4	Physiological Indicators of Trust	44
4.4.1	Pupillometry	44
4.4.2	Blood Volume Pulse, Heart Rate, and Heart Rate Variability	45
4.4.3	Electrodermal Activity	46
4.4.4	Entrainment Redux: Physio-behavioral Coupling	47
5.	Measuring Team Trust	50
5.1	The Temporal Approach	50
5.1.1	Stage 1: Initial Trust	52
5.1.2	Stage 2: Team Orientation	54
5.1.3	Stage 3: Team Performance (Trust Calibration Cycle)	56
5.1.5	Considerations for the Temporal Approach	59
5.2	Constraints to Effective Measurement	59
5.1.4	Stage 4: Experience	60
5.2.1	Team Dynamics	61
5.2.2	Individual Differences	62
5.2.3	Culture	62
5.2.4	Attitudes toward Autonomy	63
5.3	Considerations for Team Trust Measurement	64
6.	Conclusions and Path Forward	65
7.	References	68
	List of Symbols, Abbreviations, and Acronyms	93
	Distribution List	95

List of Figures

- Fig. 1 A graphical representation (Judson 2019) that demonstrates increasing levels of complexity of multiperson, multiagent human–autonomy teams. In this example, the human–autonomy teams are attempting to breach an enemy line (brown barrier). Team communication is denoted by the blue dotted lines. Frame a represents a quadcopter (reconnaissance vehicle) working with a human team to identify danger areas that dismounted Soldiers or manned vehicles should avoid. Frame b represents an unmanned aerial reconnaissance vehicle and manned ground vehicle looking for enemy threats. The information about the previously identified danger area (blue circle) has been communicated to the other human–autonomy teams. Frame c represents two robotic ground vehicles and a manned ground vehicle providing smoke-screen coverage from the previously identified enemy threats (represented by the red circles). Frame d represents two manned vehicle crews and a remote weaponized robotic ground vehicle providing support by fire (red dotted lines) for the robotic breach vehicle, while an unmanned aerial vehicle is providing additional information..... 4
- Fig. 2 The purpose of this figure is to provide a more complete understanding of how different theoretical views identifying multiple types of trust are interrelated and can impact decisions when choosing measurement techniques. 9
- Fig. 3 Closer spatial distance implies a greater sense of intimacy and trust. However, technology can mitigate distance, simulating closer proximity and allowing the technology to act as a stand-in. In this figure, a robot and a human in close physical proximity (right) are using technology (the boxes) to create a virtual space to make a physically distanced human (left) seem more subjectively close..... 31
- Fig. 4 Simplified diagram representing behavioral entrainment. Team members (a and b) can be either human or IA. Each line represents evidence of team member behavior (e.g., gestures, speech, and posture) propagating toward and back to one another. Indicators (gray and black squares) on each respective member’s timeline represent durations of a behavioral engagement during the team interaction. The center (dash-dotted) line depicts durations of overlap in a and b’s behavior. Windows around the timelines indicate different sampling periods during which synchrony is computed. Changes in synchrony patterns across these windowed segments define entrainment. 32
- Fig. 5 Diagram of face with anatomic features used for coding facial expressions from action units. Action units are scored according to intensity and can include trace (A), slight (B), marked/pronounced (C), severe/extreme (D), and maximum (E) evidence, respectively. Other emotions such as “contempt” include AU evidence from either the right (R) or left (L) side of face only..... 35

Fig. 6	Marine crew communication flow during live-fire gunnery target engagements using the Wingman platform (Schaefer et al. 2019b). The left model depicts a Marine crew in the 2018 gunnery event, and the right model depicts a different Marine crew in the 2019 event. Roles are Vehicle Commander (VC), Robotic Vehicle Gunner (RVG), and Long-Range Advance Scout Surveillance System Operator (LRAS). Graphs depict the number of communication events from each sender to each recipient, where a thicker line indicates more communication. Arrows with “Crew” as the recipient indicate communication with no specific recipient (e.g., general information for the crew and status updates to facilitate shared situation awareness). When combined with performance score data, this method revealed that the LRAS operator’s direct communication with the RVG crew member (demonstrated in the 2019 event, depicted in the right-side flow map) resulted in improved target acquisition and firing times, compared to when the LRAS operator sent more generalized information to the crew in the 2018 event (in the left-side flow map)......	38
Fig. 7	Diagram of notional multimodal connections between latent psychological states and observable physiological indicators, which may be leveraged for real-time inference of trust.....	44
Fig. 8	The life cycle of trust in human–autonomy teams. Each stage identifies whose trust can be reliably assessed, the influences on that trust, the predominate type of trust (and whether this is a state or trait at this time), and recommended assessment methods for examining trust....	52

List of Tables

Table 1	Summary of types of trust in human–autonomy teams.....	13
---------	--	----

1. Introduction

Trust is fundamental to team effectiveness and impacts a variety of team processes including information sharing, decision making, and overall team success. In high-performing teams that operate in adverse, uncertain, and high-stress environments such as military, emergency, medical, search-and-rescue, and disaster relief, an inappropriate level of trust (whether too high or too low) at a critical time can result in suboptimal interactions between teammates, leading to unsafe behaviors (De Jong and Elfring 2010; Costa et al. 2018). For example, trusting too much, or “over-trust,” can lull operators into a state of complacency leading to costly mistakes and the potential for losing human lives and expensive equipment; conversely, trusting too little, or “under trust,” can cause team members to feel the need to constantly monitor one another, creating unbalanced and inappropriate levels of workload (de Visser et al. 2020). Trust is built through training as a connected group, as well as through shared experiences over time (Fulmer and Gelfand 2012; Costa and Anderson 2017).

With the continued advancement of autonomous technologies, it is likely that future team structures will include autonomous members that function not only as tools that provide support, but as full-fledged teammates (Phillips et al. 2011). While the human performance literature is rich with methods for assessing trust in pairs of individuals (or dyads), there are gaps with respect to understanding how trust should be measured in larger teams (Feitosa et al. 2020), as well as which measures might be best suited to the complexities of teams composed of multiple humans and multiple autonomous systems. Further complicating trust measurement is the fact that team trust is a dynamic state that fluctuates over the life cycle of the team and is further influenced by the social, task, and environmental contexts (Schaefer et al. 2018a; Schaefer et al. 2019a). Nevertheless, the processes by which trust develops and propagates within these larger, heterogeneous teams remain unclear. Therefore, to enable these technologies to fully support team functioning, it is critical to understand how trust is developed, maintained, and measured in the context of human–autonomy teams. Because trust is complex, a multimodal approach to measurement is needed. Therefore, in this report we describe a conceptual toolkit developed to yield a more accurate and robust understanding of trust over time in human–autonomy teams.

Given that our future society will see a rise in heterogeneous teams consisting of multiple humans and multiple autonomous systems, it is important to quantify the trust relationships between different members of the team, as well as understand the range of implications of trust on global team dynamics. From this approach, trust between people is not the same as within human–machine partnerships, or even

machine–machine partnerships. There are different factors and features of priority for these different types of interactions and how those interactions influence the propagation of trust throughout the team. Taking all these factors into account, it is critical to first understand the capabilities and limitations of existing trust measures so they can be adapted to meet these needs, or when necessary, develop new measures to support team trust assessment. To that end, this report explores the state of the art in trust assessment techniques, including those that are novel and extend beyond subjective assessments. In Section 2 we begin by defining human–autonomy teams and outlining team trust specific to human–autonomy teams. Section 3 identifies and describes different types of trust, which differ in their development and specifically how they impact human–autonomy teaming. Section 4 breaks down the different measures of trust that have been well documented in the literature, which include subjective, communication, behavioral, and physiological indicators of team trust and how these can inform a multimodal method measurement approach for effective trust calibration in human–autonomy teams operating in dynamic environments. In the last sections, we use an illustrative case study to detail a multistep approach to measuring team trust with recommendations for modality and subsequent metric selection and identify some constraints and considerations involved in team trust measurement.

2. What Is a Human–Autonomy Team?

Before we begin our discussion of trust within human–autonomy teaming, we must examine what a team means and what is meant by autonomy. It has been posited that the foundation of human–autonomy team literature is built on that of interpersonal, or human teams (Morrow and Fiore 2012). Therefore, we extend the interpersonal definition of team as meaning “two or more [team members that] interact dynamically, interdependently, and adaptively toward a common and valued goal, objective, or mission, ... have been each assigned specific roles or functions to perform and have a limited span of membership” (Salas et al. 1992, p. 4). For the scope of this work, a human–autonomy team consists of one or more human teammates, coupled with one or more autonomous systems, or intelligent agents (IAs), that collaborate to accomplish a task or goal (Demir et al. 2019).

Recent advances in artificial intelligence have endowed autonomous systems with a greater capability for independence and interdependence. These advances effectively move technology away from roles that simply support or augment human performance in limited ways, to roles as actual team members that genuinely extend the overall team dynamics and capabilities (Phillips et al. 2011; McNeese et al. 2019). In this context, IAs are autonomous entities with the ability to observe and act on their environment, as well as conduct activities toward achieving both individual

and collective goals (Russell and Norvig 2009). IAs can be computer-based entities (i.e., embedded agents) or physical entities (i.e., embodied agents, a.k.a. robots). Both embedded and embodied agent systems can perform multiple tasks using various types and levels of autonomy, making it difficult for human teammates to understand the agent's actions or decision-making processes. Therefore, successful integration of humans and autonomous systems as team members requires each to understand the other's reasoning, actions, and intentions (Schaefer et al. 2017; Chen et al. 2018). However, developing a mutual understanding between humans and autonomy is extremely complex and requires effective communication between the autonomy and their human team member.

Part of the challenge involved in teaming humans and autonomous systems lies in the fact that team processes such as communication and coordination are fundamentally different in autonomous systems, thereby creating difficulty for both entities to develop and maintain situation awareness (Endsley 1995) and share mental models (Cannon-Bowers and Salas 2001) of the environment, task, and team. Challenges can arise from suboptimal developmental decisions (e.g., interface design and agent transparency), limitations (e.g., unable to parse natural language, understand nonverbal cues, and modality of communication), and the timing of communications. Each of these elements can impact the overall team dynamic and thus must be more fully understood to ensure proper integration of humans and autonomous systems. Further, establishing bidirectional communication with an agent, even on a one-to-one basis, can be difficult and is currently an important area of active inquiry (Marathe et al. 2018). Figure 1 provides a graphical representation of the degrees of complexity that can be present, illustrating the difficulty in understanding the interdependencies of human–autonomy teams.

Given the increased interest in adopting human–autonomy teaming beyond simple dyads, especially for high-risk operations, it is important to consider the challenges involved in developing and maintaining these types of teams. The literature regarding the mechanisms that underpin effective human team performance is rather extensive (DeChurch and Mesmer-Magnus 2010), though applying these concepts directly to human–autonomy teams is not reasonable in the near-term because the inherent differences between humans and IAs are not yet fully understood. It is essential to recognize that IAs behave, reason, and communicate differently than people and therefore it is necessary to understand how these differences affect the ability of team members to coordinate, communicate, and understand one another, as well as fundamentally change the dynamics of teamwork and the development of critical processes, such as trust within human–autonomy teams.

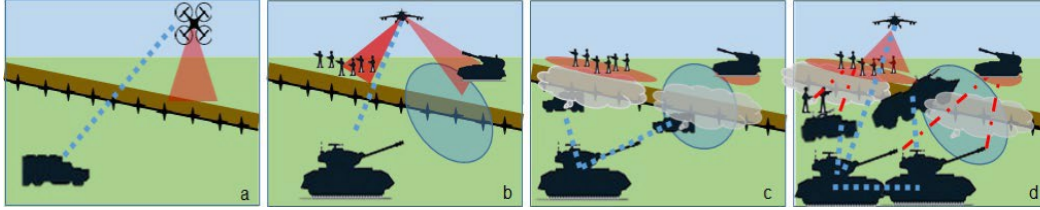


Fig. 1 A graphical representation (Judson 2019) that demonstrates increasing levels of complexity of multiperson, multiagent human–autonomy teams. In this example, the human–autonomy teams are attempting to breach an enemy line (brown barrier). Team communication is denoted by the blue dotted lines. Frame a represents a quadcopter (reconnaissance vehicle) working with a human team to identify danger areas that dismounted Soldiers or manned vehicles should avoid. Frame b represents an unmanned aerial reconnaissance vehicle and manned ground vehicle looking for enemy threats. The information about the previously identified danger area (blue circle) has been communicated to the other human–autonomy teams. Frame c represents two robotic ground vehicles and a manned ground vehicle providing smoke-screen coverage from the previously identified enemy threats (represented by the red circles). Frame d represents two manned vehicle crews and a remote weaponized robotic ground vehicle providing support by fire (red dotted lines) for the robotic breach vehicle, while an unmanned aerial vehicle is providing additional information.

2.1 Team Trust (Human–Autonomy Teams)

Human–autonomy team trust builds on the general team perspective that “team members’ positive expectations about each other’s competence and motivations, as well as a shared acceptance of vulnerability based on the assumption that teammates will act in the best interest of the team” (Grossman and Feitosa 2018, p. 396). Trust, thereby, plays a significant role in how team members interact and work together, influencing a variety of team processes and emergent states that are crucial for team effectiveness, such as task and individual team member monitoring, information sharing, and supporting team members when they need assistance (Grossman and Feitosa 2018), to name a few. Further, trust is emergent and dynamic in that, once established, it fluctuates based on a variety of factors including expectations and observations of individual and team behaviors in varied operational circumstances (Drnec and Metcalfe 2016). Given that constructs of team trust are only beginning to emerge in the relevant literature, the next section will discuss the theoretical contributions that have been laid out in the domains of interpersonal trust, human–automation trust, and information-based trust

2.1.1 Implications of Interpersonal Trust for Human–Autonomy Teams

Interpersonal trust is commonly viewed as a person’s willingness to accept being vulnerable to another person based on the expectation that their intentions, motivations, and behaviors, will be beneficial, and not harmful (Mayer et al. 1995; Rousseau et al. 1998; De Jong and Elfring 2010). Willingness to be vulnerable is a

critical component of human–autonomy team trust as people typically have varied sets of expectations for and observations of other humans and technologies across several contexts, both in and out of the relevant task context. For example, in military, search and rescue, and emergency response teams, crews of human team members often coordinate with unmanned, embodied IAs, thereby entrusting these agents with critical tasks when lives are at stake. Moreover, these crews often contain nuanced hierarchical human team elements (e.g., structural arrangement, communication chains, and so on) that directly impact the dynamics of the larger human–autonomy team.

The implications of having an interpersonal dimension of team trust are important to consider for developing human–autonomy team trust measures because individual propensities and perceptions of trustworthiness can lead to behavioral patterns that reflect a shared climate among the human team (Costa and Anderson 2011). Interpersonal trust encourages commitment to a team or organization, contributes to greater levels of member performance, and facilitates more-effective interactions while increasing understanding (i.e., reduces interactional uncertainty), as it often enables individuals to anticipate, predict, and avoid threatening or stressful situations that are characterized by high levels of risk, vulnerability, and uncertainty (Atkinson et al. 2012). Therefore, it is critical to understand which aspects of the interpersonal trust literature are applicable for measuring trust in human–autonomy teams (i.e., rather than the human crew members alone). This is not a unique suggestion as, for example, Thompson and Gillan (2010) suggest that “drivers of interpersonal trust can drive human–robot trust” (p. 78); and as such are likely to be critical influences on trust in complex, heterogeneous human–autonomy teams.

2.1.2 Implications of Human–Automation Trust for Human–Autonomy Teams

Like the interpersonal domain, the extant human–automation literature identifies several features of mutual interactions that critically affect our understanding of performance dynamics in human–autonomy teams. Therefore, they constitute important considerations for our trust measurement goals. Trust is essential as a facilitator of effective human–autonomy team performance as it enables the communication, collaboration, and coordination that underlies the teamwork needed to advance the team’s goals (Jian et al. 2000; Hoff and Bashir 2015; Schaefer et al. 2016). Trust in human–autonomy teams is similar to interpersonal trust, particularly for providing a baseline for interaction and engagement (whether with a person or an IA). As such, there are critical dimensions to consider when studying

and measuring the dynamics and influence of trust during human–autonomy teamwork.

First, in the context of human–automation teams, trust is rarely, if ever, studied in a manner divorced from the *overt behaviors* that are enacted based on the trust that the human holds (we denote these behaviors as “trust outcomes”). Even for IAs like threat alert and target identification systems, the human decision to comply (or not) with the IA recommendations is usually followed by a behavioral outcome that has real consequences in the physical world. The regularity of observing these trust outcomes has been pervasive. In a classic paper, Parasuraman and Riley (1997) characterized automation usage behaviors as falling into one of four categories: appropriate use, misuse, disuse, and abuse. While the Parasuraman and Riley framework has been very influential for understanding the variability in trust outcomes, the notion of differentiating “appropriate” from “inappropriate” implies a degree of subjectivity that can make measurement difficult. Compliance (i.e., accepting a recommendation from an IA) and reliance (i.e., opting to use available automation to manage a continuous control function) are simple and objectively defined behaviors based on observable events. Assessing behaviors like compliance and reliance may enable precise measurement and quantitative modeling of things like a person’s trust and the trust flow in human–autonomy teams (Dixon et al. 2007; Parasuraman and Manzey 2010; Meyer et al. 2014). It is important to understand these various behaviors, as they frequently are substituted as inferred trust measures despite evidence that cautions against using such measures to guide decisions (Freedy et al. 2007; Meyer et al. 2014; Chancey et al. 2017).

The human–automation trust literature emphasizes the *importance of uncertainty and risk* in the trust process, again owing to the very physical nature of this specific trust context. Understanding the roles of uncertainty and risk in team trust is especially important among and within teams that must respond adaptively to changing situations. This interdependence requires that teams establish trust among humans and autonomy alike. Trust involves decisions about the level of risk acceptable to someone based on a perceived probability of success or failure resulting from placing trust in another. The interplay between trust and risk is not unidirectional; in other words, there is arguably a reciprocal relationship between trust and risk in which they influence each other, rather than one solely serving as an antecedent and the other as solely a consequence (Kim and Koo 2016). Indeed, some authors have developed mathematical models to study how risk affects a decision to trust or not trust another (e.g., Jøsang and Presti 2004). Regardless of the perspective taken toward risk and trust, it is essential to remain mindful of the interactive and multifaceted factors that influence the manifestation of trust relevant to the scenario of interest.

Finally, the team's level of trust should be *calibrated to the situation*, where calibration refers to how well the level of trust corresponds to actual capabilities needed in the situation (Muir 1987). Too much or too little trust both have consequences for team performance and effectiveness. For example, trusting too much can lead to operator complacency, whereas trusting too little can lead to excessive monitoring of a team member or disuse of a system (De Jong et al. 2016; de Visser et al. 2020). Given that individuals continuously establish, maintain, and recalibrate their trust (de Visser et al. 2020), it is critical to measure trust over time to predict instances of inappropriate reliance so that interventions can be directed toward promoting appropriate calibration.

2.1.3 Implications of Information Trust for Human–Autonomy Teams

Information trust can be thought of as a process of attribution whereby confirmation, or convergence, from multiple sources improves the quality of the information (Wilkinson et al. 1997; Rieh and Belkin 1998; Wachbroit 2000). This concept primarily comes from cyber-security and network science whereby the nature of cyber defense requires humans to search through vast quantities of data to interpret and make judgments regarding various pieces of information presented by complex, vastly interconnected systems (Henshel et al. 2015). A common practice for addressing the types of questions surrounding information trust typically starts with initiating a “Zero Trust” protocol (Kindervag 2010), meaning that rather than assuming all actors in a network act responsibly and can be trusted, by default nothing and no one is trusted within or outside a network until it is verified. One of the benefits of Zero Trust is it provides an added layer of security, thereby decreasing the likelihood of an incident and keeping the human “in the loop” as an active team member who can confirm decisions made within the human–autonomy team. However, a concern is the potential for a substantial increase in workload, frustration, and stress for the human team member(s), which may undermine the benefits of the autonomy, and in turn, degrade team trust.

To help balance the pros and cons associated with the Zero Trust protocol, the literature describes several specific features that can be leveraged as the intelligence and capabilities of IAs continue to improve. For example, Kelton and colleagues (2007), in their integrated model of trust in information, highlight the importance of factors such as accuracy, objectivity, validity, and stability of the information, all of which impact the trustworthiness of the information. This trustworthiness, along with the disposition to information, relevance, and recommendations, feeds into the trust, confidence, and willingness to use the information and make decisions. More recently, Henshel et al. (2015) provide a trust framework from the cyber domain that identified the importance of factors such as bonding (i.e., forming an

emotional connection with the IA), the reputation of the information source, the quality of the information, the objectivity of the information, ethics including verification and validation of the information and the appropriateness (pedigree and provenance) of the source and, finally, the stability of the information over time. Together these factors help us understand and identify how trust may degrade or be appropriately calibrated within the larger team structure.

3. Team Trust Measurement for Human–Autonomy Teams

Focusing on *team trust* measurement for human–autonomy teams involves a complex set of somewhat loosely interrelated theoretical assumptions and fundamental constructs. First, while we are interested in the full scope of team trust building on interpersonal, automation, and information trust, from this point forward measurement is discussed from the human perspective as opposed to developing specific metrics that define and quantify trust from an autonomy perspective. So, from the perspective of humans, there are two main considerations for measurement that are essential to developing, maintaining, and calibrating team trust. The first consideration focuses on the premise that not all trust is created equal (see Merritt and Ilgen [2008] for similar discussion). As just discussed, trust is a complex construct that involves a large variety of influences and appears in different manifestations based on the circumstances; therefore, we take as a starting point that effective measurement must account for the operational context, as it will shift the focus onto a particular type (or set of interrelated types) of trust that are relevant to the moment. The second consideration is the construct of trust propagation. Within a team dynamic, it is critical to understand the influence of one trust relationship on the surrounding trust relationships within the team. In teams, the presence of third parties influences how trust manifests and propagates; that is, trust between two parties is subject to external influences and not simply about the dynamics of the dyad (Burt and Knez 1995; Ferrin et al. 2006). Therefore, teams of more than two represent the minimum number of interacting parties to study trust propagation and the associated demands for a well-developed method for measuring trust at both individual and team levels.

3.1 Types of Trust and Implications for Measurement

One of the major difficulties in developing effective measures of team trust is that researchers tend to use trust-based terminology interchangeably; for instance, some researchers use measures designed to assess trustworthiness as an operational definition of general trust (McKnight et al. 1998; Schaefer et al. 2021). To avoid similar confusions, we distinguish specific types of trust that we expect will have marked importance when describing team trust measurement. First, trust has been

generally conceived as being both a trait and a state (Mooradian et al. 2006). As a trait, trust has been discussed as a behavioral characteristic that is relatively stable over time and is related to one’s nature or predisposition to trust others. In the literature, this type of trust has been alternately referred to as dispositional, generalized, and/or propensity to trust—yet all three represent a type of trust that a person “carries with them,” independent of the context. State-based trust, on the other hand, is dependent on a situation or scenario and fluctuates over time. Under the premise that trust should be considered a multidimensional construct (Baker et al. 2018), we argue that to effectively measure team trust, we must consider which types of trust (e.g., trust propensity, trustworthiness, affect-based trust, cognitive-based trust, situational trust, and learned trust) (Fig. 2) will allow us to better understand and measure the trust in the team as well as its dynamic variations throughout an interaction (see Schaefer et al. 2021 for a more comprehensive review and discussion).

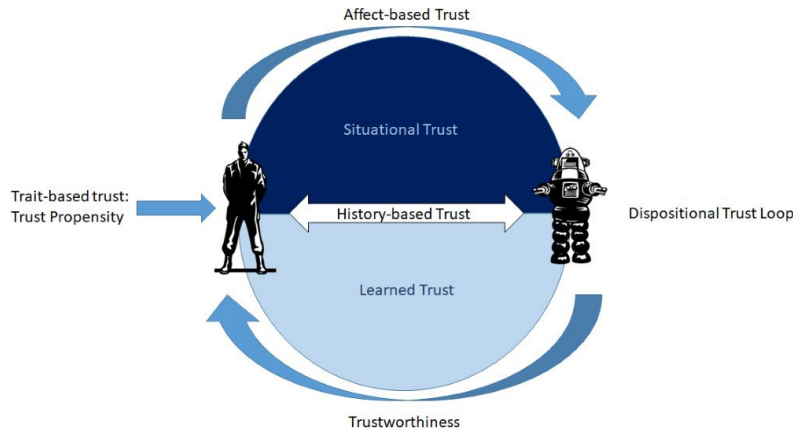


Fig. 2 The purpose of this figure is to provide a more complete understanding of how different theoretical views identifying multiple types of trust are interrelated and can impact decisions when choosing measurement techniques (Schaefer et al. 2021).

In the following sections, we will also address considerations relevant to interactions with autonomy and robots. Namely, the physical embodiment of robots and how it can affect how humans perceive and interact with them compared to non-embodied IAs. Factors such as the robot’s appearance and behavior can play a role in a human’s decision to trust it (Groom and Nass 2007; Wright et al. 2013; Pak et al. 2020); therefore, in the following sections, we also discuss when embodiment presents additional considerations for the types of trust described in the following.

3.1.1 Trust Propensity

Propensity to trust, also referred to as a general disposition toward trusting others, is a trait that remains stable over time, is unique to the individual (Rotter 1967; Mayer et al. 1995; Jarvenpaa et al. 1998), and influences the decision to enact trusting behaviors (Borum 2010; Costa et al. 2011). Trust propensity is particularly influential during the early stages of a relationship before there is enough objective evidence to determine whether someone (human or IA) should be trusted (McKnight et al. 1998; Borum 2010). As a trait, an individual's propensity to trust remains constant and is not influenced by settings or circumstances (Mayer et al. 1995; Burke et al. 2007). However, individuals differ with respect to their propensity to trust others, which is influenced by factors like personality, ethnic background, education, and socio-economic status (Costa et al. 2017). Therefore, where possible these covariates should be assessed when evaluating trust propensity. Peoples' propensity to trust robots and automation can also be assessed, which can provide valuable information before any interactions with those intelligent systems. Lee and See (2004) posited that one's predisposition to trust automation is based on prior experiences with automation and can therefore influence how future interactions with IAs will play out. Desai and colleagues (2013) found that other factors such as age and prior experience with video games could also predict an operator's tendency to trust robots. While trust propensity is traditionally measured through self-reporting, which provides important insights into baseline team dynamics, these measures are not able to capture the fluctuations in trust that occur as team members interact over time and therefore trust propensity should be measured in conjunction with state-based trust measures (see Yagoda and Gillan 2012).

In addition to trust propensity, Mayer and Davis (1999) suggest that *intention to trust* should also be considered. According to their model, trust is viewed as a willingness or intention to engage in risk-taking with someone else, based on their trustworthiness (i.e., ability, benevolence, and integrity). As such, Gill et al. (2005) report that the relationship between propensity and intention to trust depends on whether there is clear evidence to assess trustworthiness—when information about trustworthiness is clear, propensity and intention do not correlate; however, where ambiguity about trustworthiness exists, propensity and intention to trust do correlate.

3.1.2 Affective Trust

Affective trust is part of dispositional trust that a person endows toward a team member. It explains the emotional relationship between team members and is often expressed through concern for another's welfare, belief in the “intrinsic virtue” of a team

member, or the willingness to make a future emotional investment in the team (Lewis et al. 2017; Costa et al. 2018). Affective trust is an emergent attitude in which a team member makes attributions about the motives of other team members, whether human or IA (McAllister 1995; Burke et al. 2007), and may have a direct impact on comfort (e.g., familiarity and proximity), satisfaction (e.g., quality of information and dependability), confidence (e.g., attentional control/capacity), and attitudes (e.g., moods and emotions) toward autonomy (Schaefer 2013). Affective trust is critical for human–autonomy teams because IAs are designed to help alleviate high-task demands that lead to negative emotional states, poor performance, or even dangerous decision-making strategies. Affective trust can impact the emotional resilience of the team in the long term. Understanding how to measure this type of trust can provide critical insights into team members’ emotional carrying capacity, which refers to the freedom of team members to express emotions and is one of the primary means through which individual and team resilience develops (Stephens et al. 2013).

3.1.3 Perceived Trustworthiness

While affective trust is one half of what can be termed dispositional trust, or the trust a person places on another team member, perceived trustworthiness is the other half of the dispositional trust loop. From this perspective, perceived trustworthiness is developed from the interpretation of cursory information about the characteristics of the team member (human or IA) that affects the initial dispositional feelings (related to affect-based trust) toward the agent and is impacted by the environment and context in which trust occurs (Borum 2010; Costa et al. 2017). This perception can significantly influence the initial expectations of a partner. However, it is also updated over time following interactions as actual capabilities, behaviors, and features become more understood. There are several factors that affect the perceived trustworthiness of an IA and the measurement of such. Lee and See (2004) argued that an IA’s capabilities (such as level of automation), purpose, and reliability are critical to its perceived trustworthiness. Physical form is another key factor because humans make assumptions about a robot’s capabilities, intentions, and behaviors based on its appearance (Schaefer et al. 2012; Ososky 2013) and level of anthropomorphism (Dautenhahn 2007). However, these assumptions can be detrimental; a study by Robinette et al. (2016) found that people perceived an emergency evacuation-guide robot to be trustworthy because of its appearance, despite the robot’s demonstrated degree of unreliability in prior situations. Thus, the unique characteristics of robots and IAs affect perceptions of their trustworthiness. In the literature on human teams, perceived trustworthiness also includes the factors of ability, benevolence, and integrity (Mayer et al. 1995; Colquitt et al. 2007). Ability is primarily based on the trustee’s knowledge and

competence in performing expected actions. Benevolence is based on the belief that the trustee has positive intentions toward the trustor and will “do good” (Borum 2010), which can foster emotional bonds. Integrity refers to a trustee’s adhering to the principles and values deemed acceptable to the trustor, such as honesty, reliability, and fairness (Mayer et al. 1995; Costa et al. 2017; Lewis et al. 2017).

3.1.4 Cognitive-based Trust

Cognitive-based trust (also discussed as history-based trust) is established via interactions with another person or IA (McAllister 1995; Burke et al. 2007; Merritt and Ilgen 2008) and is based on existing knowledge that allows prediction of behavior given a particular task or context (Costa et al. 2018). This type of trust may provide initial trust perceptions from early interactions, but also provides critical insights into the long-term implications for team trust. As history-based trust is by its very nature continuously recalibrated from experience, interactions, and prior expectations, each successive interaction provides further information through which predictions about future behavior are made (Kramer 1999). For appropriate measurement, it is important to understand how the human team member’s experience with other team members (including IAs) influences the perceptions of competence, responsibility, predictability, dependability, and expectations for teaming (Schaefer et al. 2016). This type of trust is critical to teaming because prior to interacting with a team, people form expectations or mental models of how the members (human and IA) should behave. While interaction plays a significant role, other antecedents of trust include factors such as reliability, effective feedback, instances of errors made by the autonomy, situation awareness, and effective feedback from the autonomy, all of which can impact the operator’s perception of how to use or interact with the autonomy (Schaefer 2013).

3.1.5 Situational Trust

The concept of situational trust can be seen as a combination of both affect-based trust and a subset of features from cognitive-based trust. Specifically, Hoff and Bashir (2015) suggest this type of trust is dependent on the situation, including the environment and the human team member, allowing one to focus on a specific interaction. Features of the environment that influence trust include the type of autonomy, complexity of both the task and environment, difficulty of the task, workload and task demands of the team members, environmental risk, and uncertainty. Factors associated with the internal variation of the human team member include expertise, self-confidence, mood, affect, and attentional capacity (which according to the authors includes stress, fatigue, and boredom). Accounting for these environmental factors and variable states will help achieve more reliable measures of situational trust.

3.1.6 Learned Trust

Hoff and Bashir (2015) suggest that along with situational trust, state-based trust incorporates learned trust, which stems from training, past experiences, or even current interactions. There are three stages to learned trust: initial learned trust, dynamic learned trust (system performance), and dynamic learned trust (design features). Thus, by definition, learned trust incorporates perceived trustworthiness and a different subset of elements of cognitive-based trust that are seen in situational trust. Specifically, factors that directly affect initial learned trust include preexisting knowledge of the system, attitudes, and expectations, whereas dynamic learned trust can be impacted by system performance (e.g., reliability and validity) as well as design features such as ease of use, appearance, transparency, and feedback. These design features indirectly affect learned trust through altering the trustor's perception of the performance.

A summary of the measures just described, including why it is important for measurement, is found in Table 1.

Table 1 Summary of types of trust in human–autonomy teams

Type of trust	Why it is important for measurement
Propensity: Relatively stable trait, likelihood to trust others. Also related to intention to trust.	Sets the initial bounds for trust-based relationship
Affect: Dispositional trust that a person places on a team member (either human or autonomous)	Helps improve understanding of internal factors that influence comfort, satisfaction, confidence, and attitudes toward autonomy
Trustworthiness: Dispositional trust gained from a team member	Helps form initial expectations and mental models about the capabilities of human or autonomous team members
Cognitive/History-based: Founded on interactions with humans and/or IAs which inform perceptions of other's competence, predictability, dependability, expectations for teaming	Provides critical insights into long-term implications of team trust and how it changes over time
Situational Trust: Composed of both affect-based and a subset of features from cognitive-based trust; dependent on situation, environmental features, and characteristics of human team member	Accounting for environmental and human factors (characteristics of the human operator) leads to more reliable measures of situational trust
Learned: An operator's evaluation of a system based on training, past experiences, or current interaction; incorporates elements of perceived trustworthiness and a subset of elements of cognitive-based trust that are different from those in situational-based trust	Measuring learned trust provides an indication of how trust/reliance changes over time based on system performance and design features

3.2 Trust Propagation and its Implications for Measurement

Knowing the different types of trust and when measurement for each type should occur is a critical principle for more accurate team trust measurement. However, the team itself can impact the interpretation of those measurement techniques. Trust is an emergent construct and is influenced by an individual's cognitive, emotional, and social interactions that aggregate in complex ways to form generalized perceptions of trust shared among team members (Costa et al. 2017; De Jong and Elfring 2010). Given the interdependence of team members, it is safe to assume that attitudes and behaviors regarding trust are not independent of each other, and therefore, it follows that one member's trust in the team is expected to influence another member's trust (Fulmer and Gelfand 2012). Further, it is important to consider the interplay of these trusting relationships across all levels to include the influence of individual trust on team trust and, conversely, how team trust influences individual trust (Costa et al. 2017), as well as the degree of symmetry that exists (Delgado-Marquez et al. 2012). In other words, members may share a similar level of team trust (either high or low) and yet possess dissimilar levels of individual trust (De Jong and Dirks 2012). As a result, researchers are interested in clarifying the interplay of trust relationships that emerge within teams by considering trust asymmetries (De Jong and Dirks 2012), and the degree to which team trust consensus develops over time (Fulmer and Gelfand 2012; Costa et al. 2017). Finally, with respect to teams larger than dyads, there is a need to understand how a "third party" may influence the decision of one team member to trust other team members (Burt and Knez 1995; Delgado-Márquez et al. 2012; Ferrin et al. 2006).

This brings to light how trust propagates throughout larger human–autonomy teams. Specifically, when mapping trust in an IA, it will become necessary to determine the team trust effects across primary, secondary, and tertiary levels (Oleson et al. 2011; Schaefer et al. 2015). While much of the research has looked at the primary level (or dyadic) trust between a person and an IA, a question that follows is whether a human team member who does not have direct interaction with the IA can or does nevertheless trust it as a team member simply because they trust their human counterpart who is willing to work with the IA (and putatively is assumed to also trust it). It has been suggested that when levels of trust within a team cannot be determined from direct experience, the propagation of trust through a team will be dependent on reputation (Lewicki et al. 2006). This complexity grows as more people have different levels of direct interaction with the IA and remotely through another person or set of persons. This also affects how team trust is measured and calculated because the weight of the type of measurement may vary by team member role and degree of interactions with one or more IAs.

4. Human–Autonomy Team Trust Measurement Toolkit

This section outlines different processes for measuring team trust building on research related to self-report and communication-based indices, as well as behavioral and physiological indicators. While we have chosen not to include discussion of every possible scale or measurement technique that has ever been proposed for trust in some form, here we focus on a set of measures that, through extended team literature review, we have selected as candidates to support assessment and evaluation of team trust. Our primary criteria for selection of a given measure or technique was based on whether each instrument has been demonstrated—or at least suggested by empirical study—to be critical to human–autonomy team trust. Our goal was not only to provide a useful reference for specific measures, but also annotation with and discussion of implications as to when these measures should be used, how they should be used, and under what constraints. The goal here is to provide a toolkit for team trust measurement in human–autonomy teams.

4.1 Self-Report Measurement

Self-report or subjective scales are useful for measuring processes or constructs that are not directly observable such as operator workload, fatigue, annoyance, and comfort (Grossman and Feitosa 2018). However, self-report scales are limited in that they 1) are often taken after someone performs a task, 2) require the person to retrospectively recall how they felt at a specific moment, and 3) may reflect personal bias (Thurman et al. 2018). That said, with respect to trust, subjective measures can be effective for assessing initial trust levels or a user’s propensity to trust an IA, as well as provide spot checks of state-based trust to gain, perhaps, quick insights that tell one piece of a larger story. The inferences made through use of these scales can then be validated and further detailed through behavioral and performance data.

4.1.1 Trust Propensity Scales

Trust propensity is considered especially pertinent for early trust in the absence of information or prior knowledge of trustworthiness, but often becomes less important in long-lasting teams (Colquitt et al. 2007). Further, the propensity to trust another may also apply to human–IA teams, as the tendency of people to interact with technology or autonomy appears to constitute a stable trait or attitude that affects the decision to trust or distrust in general (Merritt and Ilgen 2008; Jessup 2018; Schneider et al. 2017).

Interpersonal Trust Scale (Rotter 1967, 1971)

One of the most widely used and validated trait-based scales is the Interpersonal Trust Scale (ITS), which measures an individual's general propensity to trust others. Respondents rate their level of agreement on a variety of statements regarding their attitudes toward groups of people and society in general. As an individual's propensity to trust is considered a relatively stable trait, the scale is typically administered one time, prior to an experiment, and outcomes would not be expected change over time.

Collection Criteria

- Participants rate 25 statements using the 5-point Likert scale.
- Scale is used for adults, typically prior to an interaction. Should only be given once per team member.

Scoring Method

- Reverse code items 1–5, 7, 9, 10, 11, 13, 15, 19, 24, then compute sum as total score.
- Scores will range from 25–125, higher scores indicate greater Interpersonal Trust.

Interpretation of Findings

- Provides indicator of a person's general attitude toward others.

Propensity to Trust Scale (Mayer et al. 1995; Mayer and Davis 1999)

The Propensity to Trust Scale consists of eight general statements in which participants rate how likely they are to trust specific groups of people. Sample items from the scale include “One should be very cautious with strangers” and “Most salespeople are honest in describing their products.” Mayer and Davis (1999) provide four additional questions to measure Intention to Trust, which has been shown to be related to propensity (see Gill et al. 2005).

Collection Criteria

- Participants rate eight general statements using a scale from 1 to 5 (disagree strongly to agree strongly).
- Administered once, prior to task or interaction, and scores relatively stable over time.

Scoring Method

- No reverse scoring; sum items and divide by the number of items to arrive at an overall score. Higher scores indicate a greater likelihood to trust.

Interpretation of Findings

- Provides indicator of an individual's general tendency to trust others.

Additional Considerations

- Provides scale to measure Trust Intention and Trustworthiness (see Mayer and Davis 1999; Gill et al. 2005).

Propensity to Trust Survey (Evans and Revelle 2008)

An alternative to Rotter's scale is the Propensity to Trust Survey (PTS). In contrast to Rotter's scale, the PTS measures both an individual's general propensity to trust others (e.g., avoid conflict), and an individual's propensity for trustworthiness (e.g., anticipate the needs of others or value cooperation). The PTS scale was validated against an investment game (Gutzwiller et al. 2019) and for use with social robotics (Langer et al. 2019; Stuck 2017). By taking this multifaceted approach to assessing trust propensity, it extends earlier efforts and may help to better support team dynamics.

Collection Criteria

- Scale consists of 21 items that measure general propensity to trust (7 items) and an individual's propensity to be trustworthy (14 items).
- Participants rate how accurately each statement pertains to them using a scale from 1 to 6 (very inaccurate to very accurate).
- Administered once, prior to task or interaction, scores relatively stable over time.

Scoring Method

- Two subscales can be scored: Trust and Trustworthiness.
- To compute a Trust score, reverse code items 4, 5, 8, 10, 11, 17, and 18 and calculate average. A higher score indicates a person is more likely to trust others.
- To compute a Trustworthiness score, calculate the average of the remaining items; no reverse coding is needed. A higher score indicates a greater degree of perceived trustworthiness.

Interpretation of Findings

- Provides indicator of an individual's general tendency to trust others and how trustworthy they perceive themselves to be.

Complacency Potential Rating Scale (Singh et al. 1993)

Complacency is an important aspect of human–autonomy teaming that we believe should be measured along with trust propensity, trustworthiness, and the like, extending beyond an individual's rating of how trusting he/she is, and enables us to examine how or whether that person's likelihood to trust technology, or in this case autonomy, is likely to manifest in their decisions regarding whether and when to use it. One early scale of direct relevance to this question is the Complacency Potential Rating Scale (CPRS; Singh et al. 1993). The CPRS references several specific types of automation and contains five main factors: general automation, confidence-related complacency, reliance-related complacency, trust-related complacency, and safety-related complacency (Singh et al. 1993). The measure has a high internal consistency ($r > .98$) and test-retest reliability ($r = .90$). Merritt and Ilgen (2008) more recently revised the scale to a 12-item scale to measure the propensity to trust machines. They found that this scale is positively related to initial trustworthiness, but not posttask trustworthiness nor automation use. However, a key concern with the current status of the original and revised scale is that some of the items are out of date and may need to be updated before it can provide accurate results.

Collection criteria

- Scale administered before task; however, attitudes develop during initial exposure to automation and become reinforced in future interactions.
- Participants rate 20 items pertaining to specific types of automation using a 5-point scale (strongly disagree to strongly agree).

Scoring Method

- Reverse code items 7 and 8. Total score is obtained by adding scores for items 1–16 (exclude 4 filler items).

Interpretation of Findings

- Range of scores can vary from 16 (low complacency potential) to 80 (high complacency).

Additional Criteria

- Scale references several specific types of automation (e.g., ATMs, VCRs, and cruise control) some of which are outdated, which may limit its

reliability. To correct this, Jessup (2018) suggests using generic referents such as “robot,” “technology,” or “autonomous system.”

Propensity to Trust Technology Scale (Schneider et al. 2017)

To address some of these limitations of the CPRS, Schneider and colleagues (2017) developed the Propensity to Trust in Technology (PTT) Scale, which measures a person’s tendency to trust technology by looking at the relatively stable characteristics of the individual, attitudes toward technology, and likelihood to collaborate with technology. A sample scale item is “I think it’s a good idea to rely on technology for help.” (Schneider et al. 2017). This scale was built on the trust in automation theory (Lee and See 2004), but in its original form had low reliability. Jessup (2018) suggests that adapting the scale items from “technology” to “automated agent” may correct this limitation.

Collection Criteria

- Participants use a 5-point Likert scale (strongly disagree to strongly agree) to rate six items that measure the general propensity to trust technology.

Scoring Method

- Reverse code item 4 and calculate average score.

Interpretation of Findings

- Measures attitudes toward technology and likelihood to collaborate with technology. Higher scores indicate a higher propensity to trust technology.

Additional Criteria

- Can modify scale term “technology” to fit system being evaluated.

4.1.2 Trust in Autonomy

Self-report scales for measuring state-based trust should be chosen to best fit the needs of the team interactions. There are several trust scales that have been created over time. The following discussion addresses a subset of those scales that we believe are sufficiently established as potentially explaining and/or understanding particular aspects of team trust. The selection of these scales amounts to assessing three key elements of trust in autonomy: perceptions of system trustworthiness, general trust in autonomy, and explainability.

Integrated Model of Trust Scale (Muir and Moray 1996)

One of the earlier trust scales addressing system trustworthiness is the Integrated Model of Trust (IMT). The IMT scale addresses the core elements of trust that

provide a judgment on system behavior. A positive aspect of this scale is that it is relatively short (i.e., 8 items total) so it provides a rapid assessment on a 0–100 visual analogue scale. Often, researchers use the first five items: competence, predictability, dependability, responsibility, and reliability to assess system trustworthiness. A limitation of this scale is that the items use the generic term “system,” which may confuse some respondents since different users possess different mental models about what the term “system” means, making interpretation of the findings difficult.

Collection Criteria

- Administered AFTER a participant interacts with a system or IA; can be administered multiple times.
- Participants rate 8 items using a visual analogue scale of 0–100. First five items typically assess system competence, predictability, dependability, responsibility, and reliability over time, with the remaining three items assessing participants’ perception of trust in system components and an overall rating of trust.

Scoring Method

- Each item is scored from 0–100 and average score computed as an overall measure of system trustworthiness.

Interpretation of Findings

- Higher scores indicate a higher level of system trustworthiness.

Additional Considerations

- Can modify term “system” to specifically address the type of system or technology being evaluated.

Human–Robot Interaction Trust Scale (Yagoda and Gillan 2012)

Extending beyond the underlying principles of the IMT, Yagoda and Gillan (2012) developed the Human–Robot Interaction (HRI) Trust Scale. While this scale originally started out looking at team trust, the finalized version resulted in a system trustworthiness scale. However, the HRI Trust Scale addresses one of the limitations of IMT by directly specifying the dimensions of the system including aspects like the user interface, sensor data, navigation capabilities, signal or bandwidth, remote information processing, automation level, and type of control. This 37-item scale uses a 7-point rating scale (strongly disagree to strongly agree) with an N/A option. The limitations of this scale are a lack of clarity in how it is scored, and how it accounts for variance between robotic systems in variable

degrees of uncertainty and risk (i.e., are certain dimensions of higher weighted importance at different times).

Collection Criteria

- Administered AFTER interaction with system or IA; can be administered multiple times.
- 37-item scale; participants rate each item using a 7-point scale (strongly disagree to strongly agree).
- Measures five dimensions of HRI: team configuration, team process, context, task, and system.

Scoring Method

- Overall scores are calculated by summing scores and dividing by number of items. Individual items can also be analyzed. There remains a lack of clarity in how scale is scored.

Interpretation of Findings

- Trust perceptions specific to human–robot interaction based on the six dimensions.

Additional Considerations

- Designed to be administered with a validated interpersonal trust scale (Rotter 1967, 1971) to account for individual differences.
- Scale is adaptable to fit experiment needs; for evaluating the team configuration dimension, a team configuration questionnaire can be given prior to the HRI scale to clearly define roles of each individual to prevent role confusion when completing the HRI trust scale.

System Trustworthiness Scale (Schaefer et al. 2012)

An additional option to assess trustworthiness is the System Trustworthiness Scale. It was developed to conduct a rapid, high-level interpretation of system trustworthiness from the key items that are important across the broad range of robotics including likelihood to use the system, perceived trustworthiness, perceived level of automation, perceived intelligence, and overall trust. This 5-item assessment uses a 7-point Likert scale and has been validated when assessing trust in autonomous mobility aids (Perelman et al. 2020), driverless vehicles (Schaefer et al. 2018b), and autonomous combat vehicles (Schaefer et al. 2019b; Milner et al. 2020) for highly intelligent and autonomy-enabled systems.

Collection Criteria

- Administered before and/or after interacting with system or IA to track changes in trustworthiness.
- Participants assess five items related to the trustworthiness of robotic systems using a 7-point Likert scale.

Scoring Method

- Each item can be looked at independently or items can be summed to get a global trustworthiness score.

Interpretation of Findings

- Allows for comparisons between people or conditions.
- To better interpret global trustworthiness, scores can be transformed into a 0–100 scale by subtracting 5 and multiplying the summed score by 3.33.

Additional Considerations

- Five scale items are shown to predict system trustworthiness across all robotic domains (Schaefer 2013).

Checklist for Trust in People and Automation (Jian et al. 2000)

One of the most widely used general “trust in automation” scales, the Checklist for Trust in People and Automation, is a 12-item scale that serves as a global measure of trust in automation; in it, participants rate their level of agreement with statements about their trust and overall impression of technology using a 7-point scale that is anchored on the extremes with the wording “not at all” and “extremely.” Similar to the IMT, this scale has a major limitation in that each item includes the word “system.” While this provides a more global use of the scale, it makes it difficult to interpret the findings as a “system” becomes an IA. People have very different mental models of what should be included in the system, and thus that perception can provide drastically different results (see Ososky et al. 2013 for implications of mental models on trust perceptions). Further, the items on this scale are highly beneficial for assessing automation-based systems but are not as accurate when assessing trust in intelligent, autonomous systems.

Collection Criteria

- Administered BEFORE as a measure of generalized attitudes toward trusting autonomy (Jian et al. 2000).

- Administered AFTER interacting with a system or IA as a measure of trust in a system; can be administered multiple times (Kohn et al. 2021; Schaefer 2013).
- Participants rate 12 items about their trust and overall impression of technology using a 7-point scale (not at all to extremely).

Scoring Method

- Reverse code items 1–5 and compute average as overall trust score. Recommended to transform scores into a 0–100 scale to better explain the findings.

Interpretation of Findings

- Can be separated into trust and distrust.
- Can be interpreted as a global trust score.

Additional Criteria

- Scale terminology can be adapted to experiment; for example, changing “system” to “robot.”

Trust Perception Scale (Schaefer 2013)

The Trust Perception Scale-HRI was developed to address trust in embodied IAs across all domains of robotics and for various levels of environmental risk and uncertainty. However, to address these specific needs more fully, for IAs the scale consists of a list of 40 items composed of a single word or short phrase that describes a perception of the robot’s capabilities, behaviors, and decision-making processes. It uses a scale from 0% to 100% of the time. A reduced, 14-item scale can be used for rapid assessment of trust over time, or when assessing trust over multiple trials or with time restrictions. However, the subscale is specific to functional capabilities of the robot and thus parallels the findings in the Checklist for Trust in People and Automation developed by Jian and colleagues (see Schaefer [2013] for discussion), though excluding a critical component of the trust explanation. Therefore, this subscale may not account for changes in trust related to robot form or features. An advantage of this scale is that it is not human role-specific and therefore can easily be used by the operator, supervisor, mechanic, peer, or bystander. This scale has also been externally validated for effectively measuring trust in both simulated and live experiments (Schaefer et al. 2015).

Collection Criteria

- Administered before (pre) and after (post) interaction with system or IA.
- Options: original 40-items scale or reduced 14-item scale.
- Participants rate scale items using a scale of 0% to 100%, which indicates the percentage of time a robot exhibited specific behaviors or characteristics.

Scoring Method

- Reverse code items denoted with an R in the scale. Items are summed and divided by number of items.

Interpretation of Findings

- Provides overall percentage of trust score. A change score can also be computed (pre–post) to track changes that occur due to individual differences, experimental conditions, or context.

Additional Criteria

- Full scale and description of development in Schaefer (2013).

Draper Trust Scales (Jackson et al. 2016)

The Draper Trust Scale is an instrument that can address the third critical element of trust in autonomy: explainability. This scale provides a framework for designing and evaluating operator trust in unmanned systems using a set of seven “Heuristics for Trusted Autonomy” that reflect characteristics of trusted systems: 1) visibility of current system behavior – what it’s doing now; 2) visibility of probable system behavior – explain what it will do in the future; 3) accessibility of system rationale – explain how decisions are made; 4) visibility of system capabilities and limitations – explain capabilities and limitations; 5) awareness of latency and delays inform user when behavior/communication delayed; 6) transparency of failure – inform user when there is a failure, off-normal conditions, and ramifications of failure on mission or performance; and 7) fit with users and operations – good match for training, expertise, and role. Each scale is rated using a 5-point Likert scale with endpoints that reflect where appropriate trust is most or least supported. Although the methods and scales are still early in development, they show promise for promoting both operator adoption and reliance on autonomous systems (Jackson et al. 2016).

Collection Criteria

- Administered AFTER interacting with a system or IA. Can be administered multiple times.
- All items are scored on a 5-point scale.

Scoring Method

- Each item scored on a scale from 1 to 5 and can either be summed or averaged for an overall score, or each scale can be evaluated separately depending on the system being evaluated and research questions. More scoring details can be found in Jackson et al. (2016).

Interpretation of Findings

- Provides an indicator of how well each scale supports or degrades operator trust.

Additional Considerations

- Provides method for a comprehensive system design evaluation (see Jackson et al. 2016).
- Initial development focused on unmanned aerial vehicles, currently extending to other domains.

4.2 Behavioral Indicators

While trust can be measured using subjective experience of internal states and judgments, objective observations of behavior are another avenue of assessment that provides additional trust-based insights. When working with IAs, humans behave in ways that reflect the extent to which they trust that agent; in response to these behaviors, agents provide algorithmic assurances—agent actions that are designed to influence human trust in the agent—which in turn affects future human behavior, continuing the trust cycle (Israelsen and Ahmed 2019). Such trust-related behaviors impact team performance and have been linked to team constructs, and thus they provide an additional layer of explanation above and beyond standard performance ratings and self-reported perceptions. Previous research highlighted several potential measures related to attentional monitoring, proximity, behavioral synchrony, entrainment, reliance and compliance, and facial analyses, explained in further detail as follows.

4.2.1 Attentional Monitoring

Trust between team members has been shown to be inversely related to the extent to which they monitor each other's behavior (Langfred 2004; Chen and Barnes 2012). Attentional monitoring can be used to assess trust in a system. Eye tracking analysis has been used to assess gaze patterns that have been linked to trust (Merritt et al. 2013; Hergeth et al. 2016). When a human operator is monitoring a system, less frequent monitoring, and hence fewer gaze fixations, may be an indication of higher trust because the human is more reliant on that system (Merritt et al. 2013; Hergeth et al. 2016). The same behavior (fewer fixations), however, could also indicate an onset of complacency; therefore, inclusion of measures to help differentiate between attributions (i.e., trust vs. complacency) may be useful. When working with lower reliability automation, operators tend to have more active monitoring of the automated tasking, as evidenced by more and longer fixations, longer scan paths, and more frequent "visits" (Bagheri and Jamieson 2004; Hergeth et al. 2016; Lu and Sarter 2019). In addition, gaze can be an indicator of trust-based decision-making over time (Lu and Sarter 2019).

Gaze findings may also be indicative of other variables. Repeated gaze fixed on an area of interest would imply that the subject finds the area useful but may also indicate that information is difficult to understand (Just and Carpenter 1976; Poole and Ball 2006). Further, gaze has been marked as an estimate of the level of cognitive effort being expended as information is being processed (Marshall 2007). Patterns and targets of gaze fixation may also signal the direction of an impending decision (Poole and Ball 2006; Glaholt and Reingold 2011).

Collection Criteria: Eye Gaze Position

- Imaging-based equipment capable of measuring eye orientation (e.g., Tobii, SmartEye, Gazepoint).
- Electrical-potential-based equipment capable of measuring relative movements (electrooculography).

Analysis Method

- Frequency and duration gaze fixations on the agent/display while performing a concurrent task.

Interpretation of Findings

- Increased frequency or longer-duration gaze fixation indicates lower trust in an agent to perform a task.
- Frequent gaze fixation on an area of a display can indicate its usefulness.

- Long-duration fixations on a display can indicate that it is difficult to understand.

Additional Considerations

- Gaze findings can indicate trust, but also usefulness, difficulty understanding, cognitive effort, or an impending decision. Regard gaze findings with caution.

4.2.2 Reliance and Compliance

The allocation of authority or responsibility between the human and autonomy can change dynamically throughout an interaction and can result in a change in the human team member's level of engagement (Parasuraman and Riley 1997; Parasuraman et al. 2000). As a result of this dynamic relationship, potential measures of trust are often related to (or conflated with) the number of operator interventions in autonomy, as well as the amount of time interacting with or manually controlling a system (Spain and Bliss 2008). In the literature, the exact nature of how these intervention behaviors manifest, and how they are measured, varies based on the function of the IA. Specifically, the construct of *compliance* is most appropriate for use with IA systems that are designed to provide alarms, alerts, or other recommendations and then, successively, leave the decision to accept the system guidance and follow the recommended (or implied) course of action to the human; the decision to follow the recommendation is then recorded as an instance of compliance. Alternatively, instances of *reliance* are most observed in systems that can operate relatively freely until an operator chooses to intervene but is not necessarily designed to base continued operation on human responses to alarms or alerts. Another factor that influences a person's reliance on a system, and potentially conflates the accuracy of engagement behavior assessments of trust, is the level of task complexity. Operator confidence in their ability tends to decrease as task complexity increases, as such their dependence on other team members, including IAs, also increases (Lee and Moray 1994; Parasuraman and Riley 1997). The distinction between these two types of behaviors, compliance and reliance, should be maintained, as the factors that influence their dynamics are believed to also be distinct (Dixon et al. 2007; Meyer 2001, 2004; Kok and Soh 2020). While it may be reasonable in many cases to believe that trust is directly reflected in the decision to rely on or comply with an IA, there is also considerable reason to question whether this relationship is so direct (Drnec et al. 2016). As such, while these behaviors—compliance and reliance—can be informative as to a person's reliance on, and ultimately trust in, a system, they should not be used in the absence of other supporting evidence.

Measure Information

- Reliance exhibited when operator intervenes (or does not intervene) upon the system when it makes decisions.
- Compliance exhibited when operator acquiesces (or does not acquiesce) to systems' alerts, requests, or demands.

Collection Criteria

- Reliance has been measured through the use or disuse of the technology when operators try to complete the relevant task (Lee et al. 1994; Parasuraman and Riley 1997).
- Compliance can be measured by examining the number of times the human operator accepts or rejects the system's recommendations or decisions (de Visser et al. 2014).

Analysis Method

- Assessing the number of trusting behaviors (i.e., instances where operator relied on or complied with the system; Hoff and Bashir 2015).
- One can use the lens of signal detection theory to interpret reliance, including system use (hit), correct rejection, misuse (false alarm), and disuse (miss; Mercado et al. 2015).
- One can calculate indicators, such as response bias and sensitivity, to understand operators' responses (Wang et al. 2009).

Interpretation of Findings

- There is a correlation between automation reliance and subjective trust, with greater reliance suggesting greater subjective trust (Parasuraman and Riley 1997; Merritt et al. 2013).
- Systems' misses degrade reliance while systems' false alarms degrade compliance (Meyer 2004; Dixon and Wickens 2006; Wiczorek and Manzey 2010); reliance and compliance compared against system performance to determine appropriate trust calibration (Lee and See 2004; Meyer et al. 2014).
- Sensitivity of measures can determine operator sensitivity to system accuracy (Mercado et al. 2015) and response bias measures can determine reliance, with more liberal responses indicating greater reliance (Wang et al. 2009).

Additional Considerations

- System transparency—when systems describe their rationale for their decisions and behavior—helps operators calibrate their trust in the system (Lyons 2013).
- The ability to audit systems' decisions can increase compliance (Lewis et al. 2021).
- There is a correlation between automation reliance and subjective trust, with greater reliance suggesting greater subjective trust (Parasuraman and Riley 1997; Merritt et al. 2013).

Systems' misses degrade reliance, while systems' false alarms degrade compliance (Dixon et al. 2007); reliance and compliance can compare against system performance to determine appropriate trust calibration (Lee and See 2004).

4.2.3 Proximity

Proxemics, the unconscious requirements of personal space and how that interpersonal distance affects social behavior, is a potential indicator of human–autonomy team trust (Henkel et al. 2014; MacArthur et al. 2017). Specifically, physical proximity corresponds with psychological intimacy, such that physical closeness connotes a personal or intimate relationship (Henkel et al. 2014; Saunderson and Nejat 2019). Studies have suggested that when humans are more familiar with a robot, or consider that robot a member of their in-group, they maintain closer proximity to that robot when working with it (Haring et al. 2013; Steain et al. 2019). When implementing this paradigm, however, one should note that more recent research has focused on the effects of robots' physical proximity to humans rather than on humans' physical proximity to the robot as a measure of trust (Henkel et al. 2014; MacArthur et al. 2017). Consequently, this behavior may be better used as confirmatory indicators.

The specific implementation of an IA, and the kinds of interactions that it can have with people, may further influence the role of proximity in team trust. When it comes to proxemics, human's responses—for example, trust—to a fully autonomous, embodied system are expected to parallel responses to other humans (Deligianis et al. 2017). Given that embodied IAs are physical platforms that can contain multiple types and varying levels of autonomy, it may be difficult to discern what is managing the proximal trust-based response. Furthermore, technology, and the way that people interact with that technology, can shape the socio-spatial arrangements that people have with it and with one another (Pelikan et al. 2018). In some situations, people arrange themselves in an unusually intimate proximity to see a

screen or be seen through a camera, implying a level of trust that does not exist (O'hara et al. 2011). Technology can even mitigate spatial distance by simulating physical proximity using cameras, virtual spaces, and telepresence systems; using technology, individuals may exhibit intimacy with a human or robotic teammate despite vast distances between them (Nguyen and Kiang 2011). In more constrained situations, proximity to an embodied IA can be a useful indicator of trust (Fig. 3), but as scenarios grow more complicated—with multiple agents, multiple robots, multiple humans, and multiple spaces—the utility of proxemics as a measure of trust may become more complicated.

Measure Information

- Proximity are methods for assessing in-group behavior.

Collection Criteria

- Proximity can be assessed through observational coding of intra-team behaviors during critical interactions.

Analysis Method

- Analysis methods include distance and timing of coded behaviors.

Interpretation of Findings

- Patterns of change in proximal behaviors over time are an effective indicator of trust; further, shorter distance also coincides with increased familiarity and in-group affiliation (Haring et al. 2013; Steain et al. 2019).

Additional Considerations

- Factors that influence proximity also include individual differences, task considerations, and mission needs. Consequently, proximity findings are difficult to generalize.

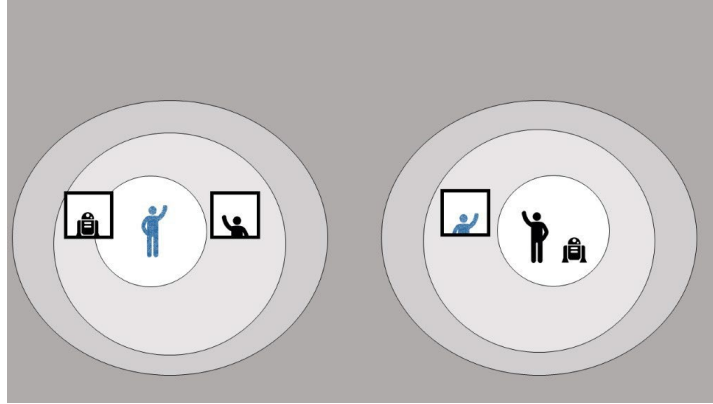


Fig. 3 Closer spatial distance implies a greater sense of intimacy and trust. However, technology can mitigate distance, simulating closer proximity and allowing the technology to act as a stand-in. In this figure, a robot and a human in close physical proximity (right) are using technology (the boxes) to create a virtual space to make a physically distanced human (left) seem more subjectively close.

4.2.4 Entrainment: Behavioral Synchrony

Behavioral synchrony, while most prominently related to trust in interpersonal teams, may provide critical behavioral measures for assessing human–autonomy team trust. Behavioral synchrony focuses on team coordination within varying levels of complexity. These mutually cooperative behaviors can include team communication, body position, and even affective cues (Strang et al. 2013). Some behaviors are explicitly controlled to be coordinated (Shockley et al. 2003), such as when an individual purposely attempts to follow the behaviors of their teammate or team leader, while others are implicit or enacted unconsciously. Additionally, such behavior is typically evident when working on a mutual, goal-directed task (Marsh et al. 2006). Presently, these shared behaviors have been referred to as entrainment (De Looze et al. 2014; Schmidt and O’Brien 1998), mimicry (Chartrand and Bargh 1999; Lakin and Chartrand 2003), and even action-matching (Bernieri and Rodenthal 1991; Henning et al. 2001); however, the specific term typically used refers to the underlying theoretical cause for the phenomenon. Theoretically, entrainment can be explained as the spontaneous interpersonal coupling that occurs because the behavior of one or more individuals is affected by another’s behavior in a closed loop system (Fig. 4).

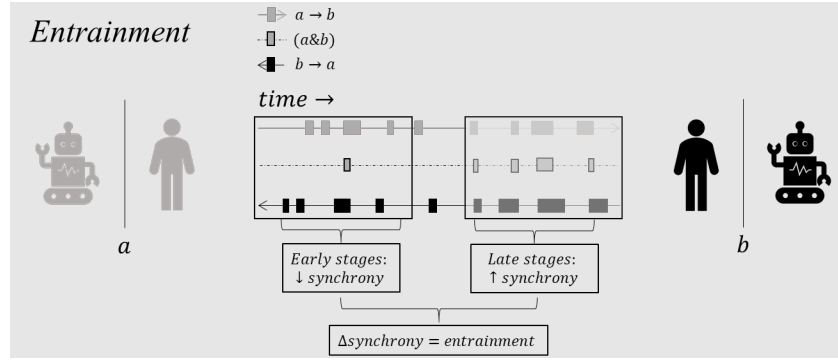


Fig. 4 Simplified diagram representing behavioral entrainment. Team members (a and b) can be either human or IA. Each line represents evidence of team member behavior (e.g., gestures, speech, and posture) propagating toward and back to one another. Indicators (gray and black squares) on each respective member’s timeline represent durations of a behavioral engagement during the team interaction. The center (dash-dotted) line depicts durations of overlap in a and b’s behavior. Windows around the timelines indicate different sampling periods during which synchrony is computed. Changes in synchrony patterns across these windowed segments define entrainment.

A large of amount of research has been dedicated to understanding interpersonal team coordination (i.e., entrainment) including studies assessing walking next to a partner in tandem (van Ulzen et al. 2008), postural sway while engaged in a mutual conversation and swinging a hand-held pendulum, (Schmidt and O’Brien 1998; Shockley et al. 2003) respectively, individual eye tracking (Dale et al. 2011) and even cardiac rhythms (Henning and Korbela 2005; Konvalinka et al. 2011). Much of the research investigating such behaviors supports the notion that increased team entrainment relates to better task outcomes (e.g., improved performance, emotional state, and task engagement (Newtson 1994; Gorman et al. 2012); however, most of the previously mentioned studies involve team members who performed identical tasks.

For measurement purposes, previous research indicates that higher entrainment may be used to predict better team performance following a role-switch (Henning and Korbela 2005). Conversely, some have argued that lower levels of entrainment may relate to better team performance when team members assume different task roles. For example, the findings of Newtson (1994) first suggested that multiperson teams who perform separate but interrelated tasks may exhibit “complementary” or asynchronous behaviors (e.g., eye contact, verbal communication, body position, and gestures). Such asynchronous behavior may result in lower levels of entrainment but may also be essential for good team performance (i.e., waiting for a command to be spoken; Sebanz and Shiffrar 2007); however, this type of outcome has only been moderately validated. This is important to consider within human–autonomy teams because in these scenarios it may be common for humans and agents to perform markedly dissimilar tasks that

will differ in their physical, temporal, or cognitive demand. As such, task roles and subsequent team performance should be considered when measuring entrainment during team interaction.

Although team coordination is a well-researched topic, the application of constructs like entrainment remains relatively novel to the area of study focused on trust in human–autonomy teams. As such, there are no methods commonly accepted for robustly indexing trust influences on behavioral synchrony in various forms. Current methods for investigating entrainment typically focus on summary statistics (e.g., average amount or frequency of a behavior), which only provides a fixed reference point about the behavior and thus does not fully capture the dynamic exchange between teammates (Russell and Norvig 2009; MacMillan et al. 2004). This approach tends to support a “random” explanation for any deviation in “average” or normal behavior; however, it is this variation in behavior that may reveal important behavioral trends such as how an individual or team is currently responding to a task in real time (Manser et al. 2008). Furthermore, patterns of interaction between team members tend to reoccur in time and are typically characterized by brief periods of spontaneous interaction followed by periods of inactivity (Gorman et al. 2012). Dynamical systems approaches, such as recurrence quantification analysis (RQA) (Eckmann et al. 1987), may present more accurate and robust measures of team trust.

Collection Criteria

- Observational coding of behaviors of interest that are performed during a task, including communication, body position, and even affective cues, entrainment, mimicry, and action-matching.

Analysis Method

- Analyzed offline to determine cross-correlations that range from 0 to 1 and emergent synchrony or via RQA.

Interpretation of Findings

- Findings suggest if the signal between individual members is more or less synchronized over time it may be indicative of greater team coordination, cohesion, and hence, trust.

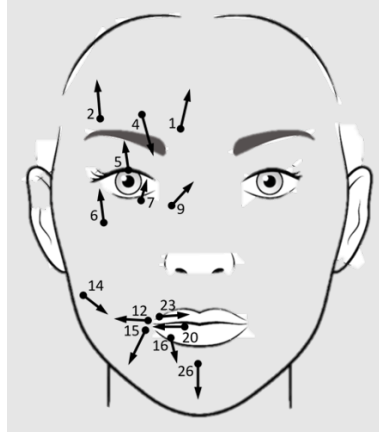
Additional Considerations

- Spontaneous interpersonal coupling that occurs when working on a mutual task; emergent similarities in behavior activity of two or more team members engaged in a cooperative behavior, which is generally spontaneous and unintended.

4.2.5 Facial Analysis

One area of emerging research has centered on analyses of emotion via changes in prominent facial features surrounding the human face, such as eyes, eyebrows, eyelids, nose, and lips. Facial expressions are only one of many indicators of human emotion, but they might be the most apparent and could be particularly important for trust development within a team (Kret 2015). Humans can produce thousands of expression variations; however, only a small set of specific facial configurations are associated with so-called “universal emotions” (Ekman 1970; Keltner et al. 2003). Though there continues to be considerable disagreement on which emotions should be considered fundamental, typical examples include emotions that express things like happiness, anger, surprise, fear, contempt, sadness, and disgust (Izard 2011). To date, collection of this data modality has been exploratory; however, we anticipate that facial expression measurements will provide corroborative support for other relevant behavioral and physiological measures that indicate changes in emotional state, trust, or team cohesion such as communication metrics, electrodermal activity (EDA), or heart rate variability (HRV).

Facial expressions can be assessed in a few different ways, such as by recording the activity of facial muscles via electrodes attached to the skin, otherwise known as facial electromyographic activity (fEMG). Muscle movements in the face create electrical impulses that are generated by respective muscle fibers during muscle contraction. An example can be seen via movement in the right/left zygomaticus, a muscle that extends from the corners of the mouth to the cheekbone. When smiling, this muscle draws the mouth up and out to outwardly produce facial movement associated with joy or happiness. Therefore, when changes in facial expressions are apparent, so too are the associated electrical impulses that result from movement. An alternative to recording facial activity is live observation and manual coding of facial activity using the Facial Action Coding System (FACS), a standardized classification system of facial expressions, which is based on anatomic features used by expert human coders (Ekman and Friesen 1976). Here, coders analyze videos of an individual’s face to determine if there are any changes in the fundamental facial components known as action units (AUs). Individual AUs correspond to specific facial muscles and are identified by a number (e.g., AU1, AU2, and so on). These AUs are then grouped together to represent universal emotions. In other words, facial expressions can be likened to “words,” while AUs are the “letters” that make up those words. For example, evidence reflecting changes in AU 6 + AU 12 is reflective of the emotion “happiness” (Fig. 5).



Emotion classification	Action units
Anger	4+5+7+23
Contempt	R12A+R14A
Disgust	9+15+16
Fear	1+2+4+5+7+20+26
Happiness	6+12
Sadness	1+4+15
Surprise	1+2+5B+26

Fig. 5 Diagram of face with anatomic features used for coding facial expressions from action units. Action units are scored according to intensity and can include trace (A), slight (B), marked/pronounced (C), severe/extreme (D), and maximum (E) evidence, respectively. Other emotions such as “contempt” include AU evidence from either the right (R) or left (L) side of face only.

Facial expressivity metrics are well-suited to measure changes in *affect-based trust*. Analyses using these features may be important to human–autonomy teams because these data (e.g., emotions, body postures, and facial expressions) can provide insights into behavioral patterns that have been linked to affiliation, empathy, and assessments of team member trustworthiness. It is vital that team members express and reciprocate appropriate facial expressions while they work together. For example, if a team failed or made a mistake in some way and one teammate smirks or laughs while the other expresses sadness or some other negative-valence expression, that exchange may degrade trust. In other words, mismatched or nonreciprocal facial expressions may lead to a decrease in cooperative behaviors (Chu et al. 2019). Conversely, if all team members exhibit appropriate expressions during critical moments in the task, then team trust should emerge. Although currently an exploratory approach, this line of research is critical because it will be necessary to develop autonomy-enabled systems that can robustly perceive and respond to emotions if human–agent teams are to be successful (Bartlett et al. 2004).

Collection Criteria

- Video recordings of the face for both a baseline and during system/team interactions.

Analysis Method

- Software analysis through OpenFace Affectiva, or Visage. Data are analyzed frame by frame and target specific muscle regions of the face (e.g., AUs) to determine the degree of evidence. AUs are then calculated and grouped to classify universal emotions (Fig. 5).

Interpretation of Findings

- Findings mark the degree to which an emotional expression is present on the human face in response to task demands or interactions with teammates.

Additional Considerations

- Nonverbal features from teammates show us proper responses; inappropriate responses (e.g., smile when team fails) could negatively impact cohesion and trust (Ekman and Friesen 1976; Kret 2015; Neubauer et al. 2020).

4.3 Communication Measures

Team communication is a rich source of information that can easily be obtained both in laboratory and field settings and, through different analysis methods, can shed light on cognition, emotion, and emerging team processes and states. Although team communication plays a vital role in driving team success (and the determinants of success), measuring team trust as a function of communication is still a new area in team research. As such, the field affords many opportunities to determine which communication analyses are most beneficial for understanding team trust or cohesion. This section describes several analytic approaches that may yield measures or proxy measures of team trust or cohesion in human–autonomy teams. For a more detailed discussion of communication measures for trust in human–autonomy teams, including implementation, considerations, and prior use cases see Baker et al. (2021).

4.3.1 Communication Rates

The rate of interaction between team members provides insight into the level of trust between them. Through repeated interactions, teammates can update their assessments of trust in their fellow teammates, while observations of these rates can be reflective of different trust dynamics. Typically, increased rates of

communication suggest increased trust (Bromiley and Cummings 1995). Beyond communication frequency, contact duration and emotional intensity also predict increased trust (Burt et al. 1995). As such, interaction histories provide meaningful insight into past and present trust relations (Kramer 1999). As an inherently social construct, exchanges between team members that are reciprocal have been shown to enhance trust, and when reciprocity is lacking, trust erodes (Deutsch 1958; Pilisuk and Skolnick 1968; Pilisuk et al. 1971; Lindsold 1978). Even within an organizational context, relationships characterized by higher trust promote greater levels of information exchange and knowledge sharing (Zaheer and Harris 2006). This is a self-perpetuating process as exchanges build relational inertia, leading to subsequent information exchanges. Trust has also been shown to enhance the rate of intra-group knowledge sharing within organizations (Staples and Webster 2008). Communication rates alone may not always reflect trust relations; however, for example, Sarker et al. (2011) found that the *type* of communication was also a reflection of a person's trust in a virtual team member. Specifically, the authors characterize communication from low-trust individuals as “unproductive or meaningless babbling” (Sarker et al. 2011, p. 299). As such, using communication to infer trust requires a multifaceted approach accounting for multiple factors, such as content, rate, and flow. A key challenge in the development of communication measures for the assessment of trust and cohesion is that much of the research into the relationships between communication and trust is presented at a conceptual level. While we discuss in this section how communication rates, flows, and content indicate trust or the absence thereof, the field has not converged on specific quantities of rate, flow, or trust-relevant content that shape trust.

4.3.2 Communication Flow

Communication flow is the assessment of how often different team members communicate with each other, as well as the evaluation of the interaction patterns they exhibit. To analyze communication flow, data are needed regarding the sender and receiver of each communication event, as well as timestamps for each communication event, if possible. One critical consideration for measurement is that the communication modality (e.g., speech, text, messaging, and so on) will affect how these data points are defined and collected. For instance, for verbal interactions, the “sender” would be defined as the speaker, the “receiver” as the intended listener, and the “timestamp” could be defined as the time the speaker began to speak. In contrast, computer-mediated interactions could be exported automatically but may be difficult to discern immediately upon receipt. A benefit to this measurement approach is that it can be applied to teams of varying sizes and compositions and used across a variety of contexts (e.g., robotic surgery, [Tiferes et al. 2016]; search and rescue, [Fischer et al. 2007]; and human–autonomy team

lethality operations, [Baker et al. 2020]), so long as there is information being sent and received between the teammates. In addition, this type of data can be used to produce flow maps (Fig. 6), providing a visualization to infer team member interactions or isolation over a longer timeframe, which may yield important indications of team trust (Hung and Gatica-Perez 2010; Tiferes et al. 2016; Baker et al. 2020).

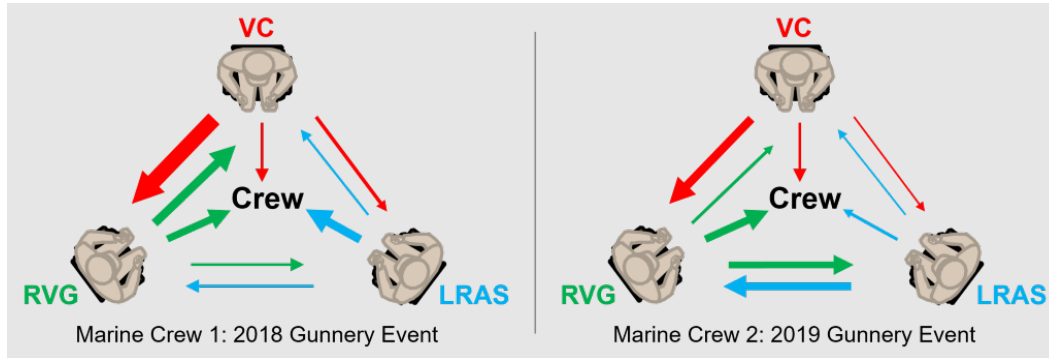


Fig. 6 Marine crew communication flow during live-fire gunnery target engagements using the Wingman platform (Schaefer et al. 2019b). The left model depicts a Marine crew in the 2018 gunnery event, and the right model depicts a different Marine crew in the 2019 event. Roles are Vehicle Commander (VC), Robotic Vehicle Gunner (RVG), and Long-Range Advance Scout Surveillance System Operator (LRAS). Graphs depict the number of communication events from each sender to each recipient, where a thicker line indicates more communication. Arrows with “Crew” as the recipient indicate communication with no specific recipient (e.g., general information for the crew and status updates to facilitate shared situation awareness). When combined with performance score data, this method revealed that the LRAS operator’s direct communication with the RVG crew member (demonstrated in the 2019 event, depicted in the right-side flow map) resulted in improved target acquisition and firing times, compared to when the LRAS operator sent more generalized information to the crew in the 2018 event (in the left-side flow map).

Collection Criteria

- Collected during interactions with human and autonomous teammates.

Scoring Method

- Sum the number of interactions between each possible sender–receiver pair.

Interpretation of Findings

- Evaluate the differences in communication flow exhibited by different sender–receiver pairs. Compare to performance outcomes to understand how more or less communication between a given pair associates with performance. Observe whether high or low flow between a given pair resulted in higher or lower trust.

Additional Considerations

- Consider looking at “baseline” teams as a comparison point for other teams’ flow patterns. For example, data from a team of experts can be analyzed to show what ideal flow patterns may look like; comparing those to patterns exhibited by a team of novices can reveal where coordination issues may lead to decreased trust.

4.3.3 Communication Content

The specific content being communicated between team members can provide deep insight into trust. For example, individuals with higher trust engage in higher levels of productive dialogue during joint problem-solving tasks (Abrams et al. 2003). Conversely, individuals who report lower trust typically check in more frequently or double check other team members’ reports to infer meaning even for inconsequential comments. This supports the idea that trust is calibrated through acts of epistemic vigilance (i.e., actively looking for inconsistencies in knowledge sources; [Yamagishi et al. 1999; Sperber et al. 2010]). Though communication content is critical to understanding interpersonal and team dynamics, sociolinguistic complexities make it difficult to know which word choices correspond to distinct constructs. Aside from explicit language (e.g., “I trust you”), the words and phrases people use that pertain to trust, such as asking questions or expressing agreement, can also stem from influences that are irrelevant to trust. Therefore, when seeking to understand team trust through communication content, it is wise to consider multiple models for how such content could reveal important trust dynamics. Here, we briefly outline two broad approaches for understanding trust via communication content: top-down approaches, which assess whether prespecified constructs or processes are present within the content, and bottom-up approaches, which infer the presence of constructs and processes based solely on word associations within the content

4.3.3.1 Top-Down Measures of Communication Content

Perhaps the most intuitive method of evaluating constructs from content are top-down approaches—computationally, such approaches require a dictionary of construct-relevant words, which are then used to quantify how often the construct is represented in the content. Developers of the Linguistic Inquiry and Word Count (LIWC) software (Pennebaker et al. 2007) have worked toward formalizing and validating several dictionaries for categories ranging from words that contain no content (i.e., function words, such as conjunctions and pronouns), to words pertaining to social processes, such as expressing positive and negative affect. Several of the dictionaries available in LIWC may help researchers understand team

trust dynamics, including sentiment (positive and negative emotions), interrogatives (i.e., question words), cognitive processing (i.e., words related to causality, insight, and certainty), drives (e.g., words pertaining to affiliation, achievement, power, reward, and risk, all of which may moderate, or contextualize, team trust relationships), and content-free function words (the basis of some linguistic similarity measures). Dictionary-based approaches offer immense flexibility and transparency because researchers can customize dictionaries for their specific needs. For instance, Frimer et al. (2014) developed a prosocial dictionary that contains 146 words and word stems, including “cooperate,” “help,” and “fair.” In this case, any use of these words during communication would be indicative of prosociality, which may in turn affect (or reflect) the trust relations between speakers. Similarly, Donohue et al. (2014) created a dictionary to assess trust and distrust using content from political speeches made during times of conflict (though this may be ineffective in team contexts due to the inherent differences between selecting words for prepared statements, such as speeches and interviews, and on-the-fly team communications). Building trust lexicons from human–autonomy team communications would be a step forward, especially for developing dictionaries to represent the factors of trust just described (e.g., trust propensity, affective trust, and situational trust) in human–autonomy interactions.

Dictionary-based approaches may also be used to quantify the implications of sociolinguistic processes (e.g., the process of grounding, which characterizes how two or more individuals form a mutual understanding of what is being discussed; [Clark and Brennan 1991]) on building team trust (Cramton 2001). Grounding is reflected in what people say—or, perhaps more accurately, what people *do not* say—in a conversation; once the specific content is established as common knowledge (e.g., “the *keys* are on the *counter*”), speakers use function words in place of specific content words as a way of increasing communication efficiency (e.g., “never mind, *they* are over *here*”; [Giles 2016]). This phenomenon led Niederhoffer and Pennebaker (2002) to develop the dictionary-based metric of Language Style Matching (LSM), which quantifies the degree to which speakers are using function word categories at similar rates. Indeed, LSM has been observed at higher rates among partners exhibiting greater levels of trust (Scissors et al. 2009; Nicholson 2017). Current evidence from human–autonomy interactions, particularly with respect to entrainment (when humans adopt the language of their autonomous counterparts; [Iio et al. 2009]), suggests that LSM would be a particularly useful measure of team trust, especially as IAs learn to entrain their own communication to match their human counterparts’ (e.g., Breazeal 2002). One potential limitation of this approach is that, while LSM may be indicative of trust, its effects on performance may also be *moderated* by levels of trust (Carmody et al. 2017), complicating the use of LSM as a linguistic measure of trust in teams.

Considering that levels of trust are dependent upon perceived certainties (Kollock 1994), researchers have also investigated evaluating trust by examining the use of epistemic adverbs (Catenaccio 2017), which qualify the certainty of a speaker's statements (e.g., probably, maybe, and definitely). To the authors' knowledge, such analysis has not been adopted in team research, though analyses of basic comprehension have shown that epistemic markers have a notable impact on the perceived plausibility of claims (Isberner et al. 2013). Other trust-relevant lexicons, such as those representing competence (e.g., Geiskkovitch et al. 2019), sentiment (e.g., Yamamoto et al. 2008; Ceron 2015; Zhang and Zhong 2019), and politeness (Linde 1988; Lam 2011) should also be evaluated in human–autonomy teams to better understand which sociolinguistic constructs relate to particular team trust dynamics. Ultimately, computing how often a construct is represented in text is made trivial by dictionary-based approaches; however, specifying which words reflect a specific aspect of team trust, for all words in a language, would be a daunting task for any researcher, though the development of better machine learning tools may improve this process.

Collection Criteria: Top-Down Measures of Communication Content

- Collected during interactions with human and autonomous teammates using speech or written language.

Scoring Method

- To assess sentiment, cognitive processes, interrogatives, prosociality, and other specific constructs, compute the proportion of observed words that match those found in a specific dictionary.
- To assess linguistic similarity between speakers, compute the average similarity between speakers across all function word categories (e.g., personal pronouns, prepositions, conjunctions, and so on). See Gonzales et al. (2010) for details.

Interpretation of Findings

- Larger numbers indicate a greater use of words from a specific dictionary.

Additional Considerations

- Proportion data can be unstable when observations are small, which can be problematic when dialogue is short or words pertaining to specific constructs are sparse, so the overall word count should not be ignored.

4.3.3.2 Bottom-up Measures of Communication Content

Top-down linguistic measures are typically static and context-independent, which may cause issues when applied to the dynamic contexts within which human–autonomy teams would perform. In contrast, models such as Latent Semantic Analysis (LSA; [Deerwester et al. 1990]) and Latent Dirichlet Allocation (LDA; [Blei et al. 2003]) take a bottom-up approach to recover the underlying semantic structure of a corpus. LSA, for instance, uses singular value decomposition to reduce a large term-document matrix into a manageable number of topics, whereas LDA (and other topic models) use a generative model to infer the underlying topics that produced the observed word and document associations. These bottom-up approaches allow researchers to understand the semantic structure of communication at lower levels (e.g., relevance between subsequent utterances) and higher levels (e.g., similarity between sources, such as two or more team members). To the extent that two speakers’ utterances are similar, one can infer that the two are speaking on similar topics. Taking this a step further, researchers have theorized that semantic similarity metrics can reveal the degree to which an individual’s ideas are incorporated into the larger group (so-called “social impact”), as well as the degree to which an individual incorporates others’ ideas (so-called “responsivity”; [Dowell et al. 2019]). While the implications for team trust are still being assessed, we would expect that team members whose contributions tend to impact the group would also be among the most trustworthy, while those members who readily incorporate others’ contributions would be among the trusting. Using metrics of social impact and responsivity as indicators of trustworthiness and trust, respectively, one could test whether contributions from trustworthy, and trusting, sources exhibit other communicative properties, such as novelty—if people recognize that their contributions are readily incorporated into the group, perhaps they would feel at liberty to propose novel solutions, with the expectation that those proposals, too, will be incorporated into the larger group. Such an approach could provide further insight into prior work demonstrating that trust allows team members to engage in task conflict (discussing different beliefs and ideas for solving problems) without allowing that conflict to affect their personal relationships (Simons and Peterson 2000). Thus, we may find the extent to which individuals’ novel ideas are incorporated, or considered for incorporation, will reflect team trust.

These bottom-up approaches are also flexible enough to perform conceptually similar tasks to the top-down dictionary-based approaches. By pre-specifying a corpus that contains semantic information on a known topic (rather than a narrowly defined dictionary of content-specific words), one could examine the similarity between the unknown source (e.g., team communications) and the known source (e.g., writings on the semantics of trust). Gefen et al. (2020) used this approach to

create a semantic space under which the linguistic features associated with measuring trust could be compared against linguistic features describing various psychological phenomena. Their resulting model could easily be used to examine communication data for any team size or composition. This type of approach has also been used to analyze the degree to which communications during a simulated mission conform to documentations regarding mission planning and execution (e.g., Martin and Foltz 2004). Although the relevance of communications to mission planning may not directly illustrate the utility of using LSA to measure trust, the flexibility demonstrated by such examples could inspire bottom-up approaches for identifying content pertaining to interpersonal trust, trust in autonomy, team trust, trust in processes, or other constructs. Further, bottom-up approaches can incorporate nonlinguistic content, such as image (e.g., Lu et al. 2010) and behavioral categorizations (e.g., Chen et al. 2019), making such approaches capable of analyzing communication content alongside contextual factors. Of course, the utility of such approaches in teams, especially with evolving IAs, is still awaiting empirical evaluation.

Collection Criteria: Bottom-Up Measures of Communication Content

- Collected during interactions with human and autonomous teammates using speech or written language.

Scoring Method

- After fitting a topic model, semantic similarity is computed as the cosine similarity between features in the high-dimensional topic space. This number is bounded between -1 and 1 or, if computed in positive space, between 0 and 1 .

Interpretation of Findings

- Numbers further from zero indicate greater similarity and numbers close to zero indicate dissimilarity.

Additional Considerations

- Quality of results can change depending on the topic model fitted to the data. If attempting to evaluate a small corpus, it may be prudent to fit a topic model to a larger publicly available corpus and then evaluate the observed data against that topic model.

4.4 Physiological Indicators of Trust

Psychological states, including those that reflect trust, do not map uniquely onto physiological indicators. Physiological response and its measurable features are driven by several interdependent internal processes and external stimuli that do not directly relate to trust states or processes. For example, pupillary response is affected by both psychological states and light levels in the ambient environment (Mathôt 2018). Similarly, measures of heart rate (HR) and EDA can vary based on relevant psychological changes as well as physical exertion or ambient temperature (Gagge et al. 1967; Kondo et al. 2000; Mikus et al. 2009). Deconflicting the signals of interest, that is, psychological states related to trust processes, from confounding effects and artifacts produced by various internal and external sources can be quite challenging (Kelsey et al. 2017). Additionally, analyzing and interpreting such signals can pose a challenge. Modern approaches to entrainment assessment have argued for the utility of nonlinear temporal structuring in team dynamics, which may yield more detailed insight (i.e., visual patterning) into such temporal signals (Marsh et al. 2006). One analysis method, RQA, has been used to assess entrainment in teams during high workload tasks. RQA is a nonlinear analysis technique that quantifies recurring states over time and represents those overlaps with approachable and intuitive visualizations called recurrence plots (Eckmann et al. 1987). We note that this is an emerging area of research and the physiology-based measures discussed in the following are also indicative of other states (stress, cognitive load, fatigue, vigilance, affect, etc.), in addition to trust-based behaviors (Fig. 7), and thus we label them as potential measures of trust along with several other interdependent latent states.

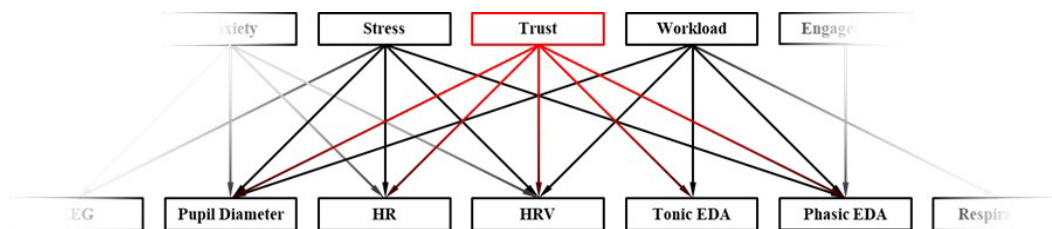


Fig. 7 Diagram of notional multimodal connections between latent psychological states and observable physiological indicators, which may be leveraged for real-time inference of trust

4.4.1 Pupillometry

The human eye appears to be one of the most salient cues that may indicate team member trustworthiness. All individuals, including infants, focus on their interaction-partner's eyes to gauge emotional signals (Farroni et al. 2002). The physical structure of the white part of the eye naturally draws attention to the middle part of the eye (e.g., the pupils) and their subsequent autonomic nervous system

changes in arousal and can thus be indicative of cognitive load, interest, surprise, and trust (Bradshaw 1967; Hess 1975; Lavin et al. 2014). For example, partners with large pupils tend to be perceived as more positive and attractive compared to partners with small pupils, who tend to be judged as distant or untrustworthy (Demos et al. 2008; Amemiya and Ohtomo 2012). Additionally, a few studies found that partner's pupil sizes can converge during interactions, although only a handful of studies have confirmed this phenomenon (Hess 1975; Harrison et al. 2006; Harrison et al. 2009).

Collection Criteria

- Imaging-based equipment capable of measuring pupil diameter is needed to collect both a baseline and task-concurrent value.

Analysis Method

- Temporal changes in pupil diameter, both transient and steady-state changes, apart from those attributable to environmental luminance.

Interpretation of Findings

- Increases in pupil diameter (not associated with environmental luminance) are generally associated with diminished trust, as well as increased stress and workload, while smaller pupils can diminish perceived trustworthiness.

Additional Considerations

- Capturing ambient environment luminance is also useful for removing artifacts associated with pupil diameter measurement.

4.4.2 Blood Volume Pulse, Heart Rate, and Heart Rate Variability

Changes in shared physiological arousal (e.g., heart rhythms) have been associated with measures of trust (Mitkidis et al. 2015) and concern for others (Konvalinka et al. 2011). Additionally, in a cooperative gaming task, Mitkidis et al. (2015) found that HR and HR synchronization increased in a trust-related condition. They also showed the level of HR synchrony within a team led to higher expectations of their partners during a cooperative gaming task and may be a marker for interpersonal trust. According to these authors, expectations are an indication of preferences and beliefs about the behavior of others; in other words, an individual may undertake an action that involves the risk of trust only if they believe this choice will be reciprocated by the other individual. Features of blood volume pulse (BVP) and EDA have also been shown to correlate with user trust in predictive decision making using an intelligent decision aide (Zhou et al. 2019).

Forgiveness, which relates to trust dynamics in that it captures the degree to which trust can be built following a transgressive act, has been shown to correlate as both a trait and state value with cardiovascular features, in the context of recounting experiences of conflict (Lawler et al. 2003). Trait level forgiveness was associated with lower levels of blood pressure, especially diastolic blood pressure. State level forgiveness was also associated with lower blood pressure levels, lower HR, and lower rate pressure product, while being unable to forgive specific offenders was related to increased levels of cardiovascular tone. Failure to forgive, especially in those who are generally less forgiving, was associated with a more sustained period of cardiovascular reactivity, and this heightened reactivity was propagated into the recovery period by these low-trait, low-state forgivers. HRV may also be used in conjunction with EDA to infer levels of workload and trust (Matthews et al. 2005; Mehler et al. 2009; Montague et al. 2014).

Collection Criteria

- Data from electrocardiographic or photoplethysmographic sensors is collected for baseline and a concurrent value

Analysis Method

- Collected rhythmic electrical or blood flow signals of the cardiovascular system are used to extract features of heart pulse, including HR, HRV, and BVP.

Interpretation of Findings

- Changes in heart rhythms from physiological arousal in human teams have been associated with measures of trust and concern for others. HR has been found to increase and synchronize between teammates in trust-related conditions of a cooperative task. Features of BVP were shown to correlate to user trust in intelligent decision aides.

Additional Considerations

- In conjunction with EDA, HRV can also be used to infer workload (Matthews et al. 2005; Mehler et al. 2009; Montague et al. 2014).

4.4.3 Electrodermal Activity

Although a direct mapping between EDA measures and trust states cannot be directly asserted, several controlled experimental paradigms have indeed linked changes in EDA to changes in states such as stress and engagement within autonomous environments (Mower et al. 2007). In this context, increases in both tonic and phasic EDA have been shown to relate to increases in anxiety and

cognitive effort (Shi et al. 2007; Zhou et al. 2015). Additionally, work by Bethel et al. (2007) found that tonic EDA measures increased with increased engagement with a robot. Furthermore, Montague et al. (2014) employed a dyadic interaction trust paradigm and found that higher individual ratings of trust in technology corresponded to more agreement in their individual EDA measures. In other words, if EDA levels of one subject were low, and both subjects were in a trust state, the second subject's EDA would be expected to be low as well. In a pilot study, Schaefer et al. (2012) found higher EDA corresponded to lower subjective trust ratings during interaction with an unreliable robot as compared to interaction with a reliable robot.

Collection Criteria

- Electrodes attached to skin or via a wrist worn sensor (such as an Empatica). Data is collected for baseline, concurrent task, A BIOPAC MP150 system with EDA amplifiers.

Analysis Method

- Skin conductance level (tonic EDA) and skin conductance responses (phasic EDA) measured during task. Specific skin conductance responses are seen shortly after an emotionally or arousing stimulus presentation.

Interpretation of Findings

- Individual activation or arousal level of individual during an emotionally stimulating task tonic EDA measures shown to increase as robot engagement increases. Synchrony between individuals can indicate a similar level of trust in a technology. Higher EDA levels have been shown to correspond to lower subjective trust ratings of an unreliable robot.

Additional Considerations

- Increases in tonic EDA are associated with lower ratings of subjective trust; however, changes in the phasic response would indicate the degree or magnitude of the response.

4.4.4 Entrainment Redux: Physio-behavioral Coupling

There has been evidence to suggest that various forms of entrainment, as introduced previously and potentially characterized through RQA (Eckmann et al. 1987), can occur in humans. These can range from body postures to facial expressions (both negative and positive emotions; [Dimberg 1982; Chartrand et al. 1999]), heartbeat (Levenson and Gottman 1983), and even verbal patterns emerge when individuals engage with familiar “in-group” members rather than “out-group” strangers

(Norscia and Palagi 2011; Hess and Fischer 2013; Reed et al. 2013). Prior work has shown that perceptions of in-group or out-group team members are important, as our tendency to trust a team is more efficient and reliable when individuals in a team are considered in-group members (Balliet et al. 2014; Cikara et al. 2014). As such, entrainment can occur in nearly every social interaction where trust is an influential factor and is a powerful effect, as it facilitates affiliation, empathy, and prosocial behavior. Like entrainment, another critical outcome of team member interaction is referred to as physiobehavioral coupling (PBC), which is defined as the spontaneous and unintended emergence of autonomic, cortical, or behavioral similarity for those who are engaged in a cooperative task (Strang et al. 2013).

PBC has related to various outcomes impacted by trust within a group and has been argued that this interaction further facilitates the development of effective interpersonal coordination within teams. Mønster et al. (2016) found that synchrony of skin conductance for members of newly formed teams was correlated with group tension and negative affect, with synchrony decreasing after the team adopted a new routine. There has been evidence to show spontaneous PBC of cardiac rhythms during cooperative tasks (Levenson and Gottman 1983; Henning et al. 2001; Henning et al. 2005; Elkins et al. 2009; Henning et al. 2009; Konvalinka et al. 2011), respiratory responses (Henning et al. 2001), brain activity (Tognoli et al. 2007; Lindenberger et al. 2009; Astolfi et al. 2010), EDA (Henning et al. 2001), and eye movements (Richardson and Dale 2005; Dale et al. 2011). PBC may present an interesting method for studying team cohesion and hence trust, given that higher PBC was found to be moderately related to better performance (Strang et al. 2013). Additionally, Behrens et al. (2019), in the context of a dyadic Prisoners' dilemma game, found a relationship between cooperation among team members and PBC, where increased HR and skin conductance synchrony between dyad partners was linked to more successful cooperation. Further, Henning et al. (2001) found the increased PBC strength of team members cardiac interbeat intervals (IBI), EDA, and lung tidal volume significantly predicted increased team performance. Past studies have also found that IBI coupling is related to successful team characteristics such as situational understanding (Konvalinka et al. 2011) and greater team awareness (Elkins et al. 2009). Finally, PBC has also been linked to oral communication and team cohesion (Barrick et al. 1998). It should be noted that many of these findings have been linked to performance, coordination, cooperation, or team cohesion, but they should be explored as a component of team trust as well.

Collection Criteria

- Various physiological sensor modalities (EDA, electrocardiogram, electroencephalography [EEG], etc.) are needed to collect both a baseline and task condition. Physiological time series data: HR, respiratory rate, EEG.

Analysis Method

- Analysis occurs offline to determine cross-correlations (which range from 0 to 1) and emergent synchrony. Cross-correlations closer to 1 indicate greater physiological entrainment.
- Data is evaluated based on the quantification of recurring patterns in a phase space using a set of standard RQA statistics: recurrence rate, determinism, trapping time, and entropy

Interpretation of Findings

- The spontaneous (physiological) interpersonal coupling that occurs when working on a mutual, goal-directed task; emergent similarities in the cortical/autonomic activity of team members engaged in a cooperative behavior that is spontaneous and unintended; can include heart beats, respiratory response, EEG, or EDA.
- For recurring sequences in the time series, these statistics indicate the frequency of recurrence, the length of the recurring pattern, and how much variation is in the length of recurrent sequences. RQA is commonly used with physiological signals and could be used to measure trust relations between two people.

Additional Considerations

- More research is needed to confirm this measure is indicative of trust in human-agent teams.
- Due to the exploratory nature of this method for inferring trust, RQA should not be used as a standalone measure but in conjunction with other established measures.

5. Measuring Team Trust

Appropriate measurement is key to successfully examining trust in teams. However, it is not as simple as choosing one of the methods from the toolkit and examining an individual out of context or extrapolating an individual's information to the entire team. Team trust is an exceedingly complex process, evolving and iterative by nature, and appropriately selecting, administering, and interpreting trust measures is crucial to further our understanding. Herein, we recommend a temporal approach for examining team trust, looking at each step in its evolution, noting important types of trust to capture and appropriate measures to use. We frame our discussion using a case study, with the intent to better ground our approach and offer concrete examples for reference. After examining each step and its implications for measurement selection and administration, we then discuss potential constraints and confounding factors that can influence measurement and interpretation.

Case Study: Consider a human driver and passenger (navigator) in an autonomous vehicle traversing an environment with varying terrain and urban development to reach multiple checkpoints. Along the way, the driver and navigator scan for items of interest (e.g., hazards), communicate with each other about such items, and log this information. Meanwhile, the autonomous vehicle communicates information to the human teammates about its own state relevant to each's role (e.g., speed and obstacle detection to the driver, route, and current location to the navigator), the task environment (e.g., road closures, detours, and traffic), and the shared task (e.g., goal locations). The team's objective is to note items of interest, avoid hazards, and reach each successive checkpoint as quickly as possible. Using this scenario, throughout the following section we illustrate the dynamic nature of trust over the course of team formation and performance and emphasize gaps in our current understanding of trust in human-autonomy teams. Ideally, this example will then be useful in informing the design of trust interventions to calibrate and support optimal team trust.

5.1 The Temporal Approach

As we conceptualize team trust as a construct that evolves and changes over time, team trust assessment must also be a continuous, temporal process. In our framework (Figure 8), we envision the team trust process can be segmented into four stages: initial trust, team formation, team performance, and experience. The Initial stage is focused on the individual's dispositional propensity to trust, which are relatively stable traits—as well as the individual's experiential trust upon completion of a task or dissolution of the team. We refer to the second stage as

Orientation, wherein the team formation begins and individuals assess other members' trustworthiness, competence, and so on. It is at the Orientation stage where team trust assessments can begin. In the Performance stage, team trust consists of a dynamic cycle as situations and risk change, an iterative process of trust development, maintenance, and calibration that continues until the team disbands. Like the Initial stage, the final stage, Experience, focuses on the individual and their history-based trust of the team. Over time, this history-based trust can influence changes in an individual's dispositional propensity to trust, closing the trust cycle. Figure 8 illustrates each of these stages for a human–autonomy team, including identifying the factors that influence trust and the appropriate assessments at each stage.

This proposed process for effective team trust measurement can be likened to Tuckman and Jensen's (1977) classic five-stage model of group development, which consists of group forming, storming, norming, performing, and adjourning stages. In their model, forming refers to orientation, where a newly formed group's members disclose information about themselves and gather information about others. The storming stage is dominated by conflicts and a lack of unity (Bonebright 2010). Norming then refers to the structural stage of group development whereby the team forms norms, roles, and expectations. The performing stage covers within-group development theory and describes how the team performs together, including healthy conflicts that strengthen bonds and do not damage relationships (Tuckman and Jensen 1977). Finally, the last stage in Tuckman and Jensen's model is the adjourning stage in which the team has reached their goal or completed the project and prepares to disband and potentially move on to other teams. While Tuckman and Jensen originally proposed this as a linear life-cycle model (Bonebright 2010), we suggest there is a cyclical nature to team development that continues until the team disbands. To understand team trust measurement, we must understand how the team develops throughout each stage.

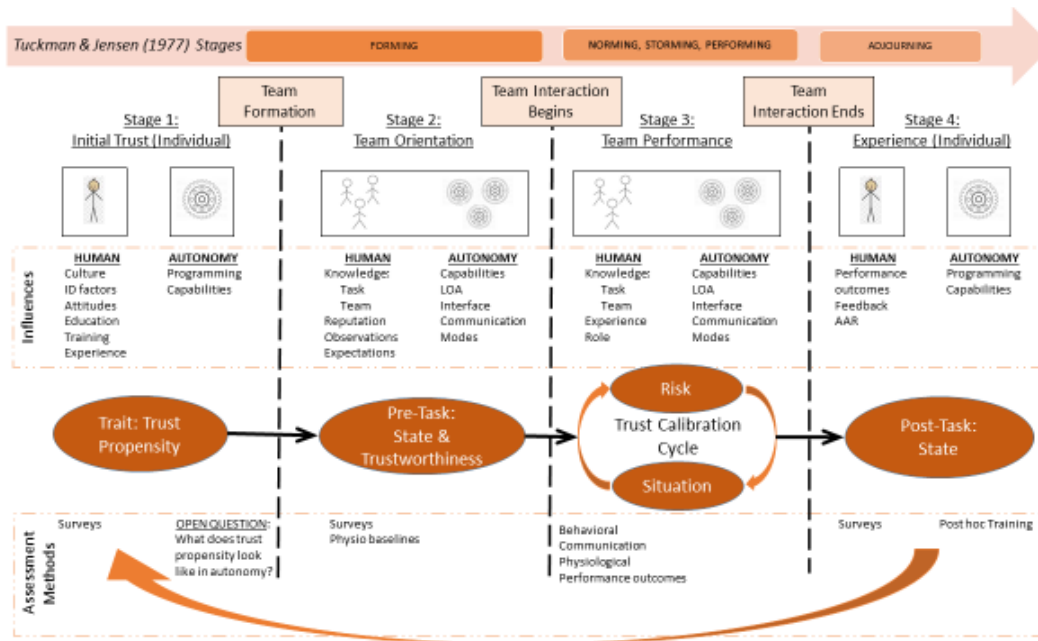


Fig. 8 The life cycle of trust in human–autonomy teams. Each stage identifies whose trust can be reliably assessed, the influences on that trust, the predominate type of trust (and whether this is a state or trait at this time), and recommended assessment methods for examining trust.

5.1.1 Stage 1: Initial Trust

There are two antecedents to the development of team trust that should be assessed prior to team formation: propensity to trust (Mayer et al. 1995; Costa et al. 2011) and intention to trust (Mayer et al. 1999). Each of these are stable traits, developed over time and influenced by such factors as culture, education, personality traits, and experience (see Section 5.2.2). These are uncorrelated and separate traits, unless the information the trustee has about the trustor is ambiguous, then these two traits are positively correlated (Gill et al. 2005). As such, the trustee’s Intention to Trust may be as critical to team trust development as their Propensity to Trust.

In human–autonomy teams, an individual’s propensity and intention to trust should be assessed toward both other people (Rotter 1967; Mayer et al. 1995; Jarvenpaa et al. 1998), and technology or autonomy (Merritt et al. 2008; Schneider et al. 2017; Jessup 2018). This baseline of trust can influence the decision to enact trusting behaviors (Borum 2010; Costa et al. 2011) and may provide information useful in understanding differences in trust at later stages. The Initial Trust stage would appear early in Tuckman and Jensen’s Forming stage, as Forming includes assessing the individual’s threshold for trusting others (Tuckman and Jensen 1977). The Initial Trust stage is critical for understanding what each team member (both human and

autonomy) brings into the team, as well as understanding differences in trust relationships among the team members.

Case Study Exemplar: As each member of the team (i.e., driver, navigator, and vehicle) arrives for training and assignment, they complete demographics questionnaires and various surveys designed to understand what influences their trusting behaviors. In terms of the vehicle (autonomy), this would mean identifying how the programming has been developed to assess other team members and situations, as well as incorporating new information (if it is adaptive or learning capable).

Suggested Assessments and Methods: In the Initial stage, assessments for the human teammates will be self-report surveys, administered individually, prior to any training or interactions with other team members. Refer to the Toolkit section (Section 4) for detailed information about each assessment tool.

- Surveys
 - Demographics information: examples include gender, age, ethnicity, background, personality, education, culture, socioeconomic status, prior training, previous experiences with autonomy, previous experience working with other teams, and task-specific experience
 - Interpersonal Trust Scale (Rotter 1967, 1971)
 - Propensity to Trust Survey (Evans and Revelle 2008)
 - Propensity to Trust Technology (Schneider et al. 2017)
 - [Intention to] Trust Survey (Mayer et al. 1999)
 - Propensity [to Trust] (Mayer et al. 1999)
 - Complacency Potential Rating Scale (Singh et al. 1993)
 - Trust in Automation Scale (Jian et al. 2000)
- Behavioral Assessments
 - None at this time
- Communication Assessments
 - None at this time
- Physiological Assessments
 - None at this time

- Autonomy Assessments
 - Programming parameters and details

5.1.2 Stage 2: Team Orientation

The Team Orientation stage includes initial team interactions, including training. Team members meet and begin to form expectations, perceptions, and mental models about one another and upcoming interactions. This stage is the basis for team trust development, maintenance, and future calibration. The team members are not yet familiar with each other. Their impressions and judgments—influenced by the factors assessed in the Initial Trust stage, including individual biases and prejudices—form the basis of the team trust foundation. Subsequently, team members evaluate their counterparts, including autonomous teammates, on their *perceived trustworthiness* and competency. Expectations developed at this stage are largely based upon superficial signals, such as gestures, tone of voice, and facial expressions. Trust in this stage is *affect-based*, and is reevaluated throughout this stage, adjusting with each subsequent interaction and observation, in an effort to reduce uncertainty related to the team and their upcoming task.

Trust development in the Team Orientation stage can also be influenced by an individual's learned trust (Hoff and Bashir 2015), so preexisting knowledge of the system and system performance, reputation, prior task or team knowledge, observed system features, and perceived team member capabilities (i.e., attitudes toward the team within a context of an upcoming task; [Schaefer 2013]) all explain some of the potential variance in trust scores. The Team Orientation stage encompasses Tuckman and Jensen's Forming stage, wherein each member is assessing the other's trustworthiness (Tuckman and Jensen 1977). However, when training is included in the Team Orientation stage, both *norming* and *storming* may begin. Norming refers to group development of team norms, roles, and expectations. Storming is dominated by conflicts and a marked lack of unity among team members (Bonebright 2010). Important to our goals is understanding that without setting appropriate expectations, researchers may misconstrue system misuse or disuse as reflecting a lack of trust when, in fact, these behaviors could be just part of the process of learning how to work with a new team and/or system during the norming and storming process.

Case Study Exemplar: The team members meet and begin their assessments of each other, the autonomy, and their task. Learning the capabilities and limitations of each team member enables the team to begin developing and calibrating trust. During Team Orientation, their initial opinions of member trustworthiness and competency can be assessed. If physiological assessments of state changes during

the task are desired, baseline measurements should be collected during this stage. While training, team members develop their interaction styles while they learn and practice the tasks. Observed behavioral data can be collected to better inform later observations that may indicate stress, workload, or trust changes. If the autonomy is adaptive or learning capable, any data it has collected during training and initial interactions with the team can be assessed.

Suggested Assessments and Methods: In the Team Orientation stage, assessments for the human teammates could be self-report surveys (both individual and team), physiological, and behavioral. Changes in the autonomy programming can be reviewed for changes after interacting with the team. Refer to the Toolkit section (Section 4) for detailed information about each assessment tool.

- Surveys
 - System Trustworthiness Scale (Schaefer et al. 2012)
 - Trust Perception Scale (Schaefer 2013)
 - Integrated Model of Trust (Muir and Moray 1996)
- Behavioral Assessments
 - Facial expressions, gestures
 - Training performance
- Communication style, tone, frequency, and content
- Physiological Assessments
 - Baseline measurements for any physio measures
 - Pupillometry
 - BVP, HR, HRV
 - EDA
 - Physiological synchrony
- Autonomy Assessments
 - Adjustments to code, tightened parameters, for example

5.1.3 Stage 3: Team Performance (Trust Calibration Cycle)

In the Team Performance stage, team members are interacting and performing their tasks as they work toward team goals. Trust development and maintenance during this stage is a continuous cycle of assessing and reassessing teammate competency, task requirements, situational changes, risk, and team performance (Savolainen and Ikonen 2016). Part of this cycle of trust includes human team members acting in ways that lead the autonomous system to react in some way, which in turn causes the human team member to recalibrate their trust, thus affecting future behavior (Israelsen and Ahmed 2019). This trust calibration cycle (Fig. 8) continues throughout the Team Performance stage. Trust in this stage is an extremely complex process and understanding that trust even more so. On the surface, interactional trust could appear to be dependent on a situation (Hoff and Bashir 2015); however, the associated risk and uncertainty define the bounds of that trust. Thus, assessment without these bounds induced by risk and uncertainty makes it difficult to interpret the true nature of trust. For example, use or disuse of autonomy may be indicative of several other factors (e.g., boredom, curiosity, and so on; [Parasuraman and Riley 1997; Drnec et al. n.d.]) and thus are not necessarily indicative of trust-based actions when risk is absent [Schaefer et al. 2016]).

Situational trust remains fluid throughout the Team Performance stage and is associated with the appraisal of trust in one's teammates with respect to specific environmental and team states or contexts. As such, it varies as a function of changes in environment, including factors such as workload, task demands, and uncertainty (all of which may change more rapidly than established trust relationships within the team). It is expected that introducing the team to situations in which trust has not yet been established (e.g., novel task environments or situations requiring unfamiliar role configurations), or changes in team member state (e.g., fatigue, boredom, and stress), will lead to greater trust volatility, due to newly forming situational trust.

While affective and learned trust evolve, over time trust becomes more cognitive/history based, as such trust should stabilize or solidify as more data and experience are gained. Team members' perceptions should gradually be less dependent on initial biases and propensities and based more on interactions and their outcomes, limiting the implications of extraneous variables on their trust. Eventually, these perceptions will be bidirectional, as autonomous systems will be developed with the capacity to build a trust model of its human teammate and act based on that model (Lahijanian and Kwiatkowska 2016; Sperrle et al. 2019). The amount of change in team trust, given an interaction, should be balanced by the length of time the team has worked together. However, trust at this stage would still be susceptible to violations and subsequent repair if the team is to be successful, but

the relative effect of those instances would be reduced as the team relationship builds, requiring more significant infractions to impact trust with similar magnitude as early experiences. Tuckman and Jensen's (1977) *storming* and *norming* stages continue into the Team Performance stage, along with the *performing* stage. The performing stage describes how the team performs together, including healthy conflict that serves to strengthen bonds and not damage relationships (Tuckman and Jensen 1977). As the team grows in familiarity and team competency, the storming and norming cycles should lessen in frequency and the emphasis move to performing.

Assessment of trust in the Team Performance stage occurs *during* team interactions. Employing physiological and behavioral measures will provide data regarding subtle or not-so-subtle changes to facial expressions and HR (linked to what is traditionally considered affect-based trust). Specifically, when members attribute setbacks to the actions of other members (including IAs), the frustration they experience will be reflected in their automatic (e.g., HRV and EDA) and volitional responses (e.g., communication and overt behaviors). These changes in trust will also be reflected in their interaction networks over time, specifically in terms of active, rather than reactive, interactions. As a result, assessing additional cognitive states, such as workload, is also pertinent. For example, workload that is either too high or too low can indicate an overload situation or underload/boredom, which may suggest (or occasionally be confused with) circumstances of over- or under-trust (de Visser et al. 2020).

Case Study Exemplar: Team orientation and training are completed and the team begins conducting its task. Initially the team is still resolving its working and interaction styles until it settles into a pattern that suits its needs and satisfies objectives. As the team moves through the environment, they continue to assess each other's reliability, competency, and performance, particularly in the contexts of task environment, situational changes, and associated risks. Example behaviors shaping trust within the team include the autonomous vehicle's avoidance of a hazard not initially noticed by the driver or the navigator's selection of a suboptimal route to reach the next checkpoint. Administering surveys would be difficult during task performance but this data type may be collected during pauses or breaks. Behavioral and physiological assessments of performance and cognitive states are most useful to understand the development, calibration, and maintenance of team trust during this stage. Multiple measures should be obtained to offer evidence either supporting or clarifying interpretations, as which cognitive states are changing is a matter of inference and interpretation. Performance data from the autonomy should be collected and assessed. If the autonomy is adaptive or learning capable, any changes to its programming should be reviewed.

Suggested Assessments and Methods: In the Team Performance stage, assessments for the human teammates could include self-report surveys (both individual and team), behavioral, and physiological. Changes in the autonomy programming can be reviewed for changes after interacting with the team. Refer to the Toolkit section (Section 4) for detailed information about each assessment tool.

- Surveys
 - Integrated Model of Trust Scale (Muir and Moray 1996)
 - Human–Robot Interaction Trust Scale (Yagoda and Gillan 2012)
 - System Trustworthiness Scale (Schaefer et al. 2012)
 - Checklist for Trust in People and Automation (Jian et al. 2000)
 - Trust Perception Scale (Schaefer 2013; Schaefer et al. 2016)
 - NASA TLX (Hart and Staveland 1988)
- Behavioral assessments
 - Proximity
 - Facial expressions, gestures, behavioral synchrony
 - Eye tracking: scan patterns, fixations, saccades
 - Task performance, response times
 - Monitoring behaviors
- Communication rate, flow, and content analysis
- Physiological measures
 - Eye tracking: pupillometry (pupil dilation)
 - BVP, HR, HRV, EDA
 - Physiological synchrony
- Autonomy assessments:
 - Adjustments to code, tightened parameters, for example
 - Observational data from times when the autonomy fails or when interactions with human team members break down because of a lack of understanding

5.1.5 Considerations for the Temporal Approach

The development stages just described were conceived to provide guidance regarding trust development in individuals and teams, implications for assessments throughout the cyclical process, and suggestions as to the types of assessments that can be collected at each stage. As demonstrated, assessment techniques will change in importance and value over the life cycle of the team. For example, an individual's trust propensity will have less of an impact on team trust later in the relationship than early on. As we explained, team trust is not a linear process with a definitive beginning and end. Overall, it is a cyclical process and within each stage there can also be cycles of trust development, changing with each interaction. Individual team members experience lasting changes to their dispositional trust due to team membership and experiences, which in turn also impacts their propensity and intention to trust in later encounters. These changes influence how they view future team membership, regardless of whether it is a new team or an established team reassembling for a specific assignment. Teams that group and disband as needed, but retain the same team members, may in subsequent groupings realize shortened Initial and Orientation stages, moving quickly into the Performance stage.

Two important aspects to team trust that we have chosen to not address in this report are trust violations and repair. Trust repair strategies may be needed if a trust violation occurs that disrupts a cycle. Trust between team members evolves over time and can strengthen or weaken in response to the actions taken by the team members. During shared experiences, trust may be violated, intentionally or unintentionally. To address the violation, the parties may attempt to repair the damaged trust to return the relationship to a trusting state. The literature suggests several approaches for repairing trust: apologies, denials, trustworthy action, or commitment to change (Kim et al. 2004; Tomlinson et al. 2004; Schweitzer et al. 2006; Lewicki 2007; Baker et al. 2018). The most effective trust repair strategy for a given scenario depends on several factors, such as the nature of the violation (e.g., stemming from competence or integrity) and prior experiences between the involved parties. In any case, the process of repairing damaged trust can be considered important to trust development between two or more parties because it can reveal information about each party's intentions, attitudes, beliefs, and affective states.

5.2 Constraints to Effective Measurement

With any measurement, there are additional variables or factors that should be considered when interpreting and explaining the measurement outputs. The purpose of this report is to focus specifically on trust due to the importance trust emergence has in human and human-autonomy teaming. However, we also acknowledge that

other latent states of interest (i.e., stress, workload, cohesion, and so on) can also co-occur along with trust during team interactions. This section identifies some of the critical factors that may affect how trust scores can and should be interpreted. These include team dynamics, individual differences, culture, and attitudes toward autonomy.

5.1.4 Stage 4: Experience

The Experience stage occurs when team interaction ends and addresses the long-term implications of team dynamics on team and individual trust. This late-stage trust is based on the individual's experience with the team, their anxiety over separation, and disengagement. Assessment of trust in the Experience stage can encompass both team trust (early) or individual dispositional trust (enduring). After-action reports and other subjective surveys can be used to examine the individual's views of the team as well as any changes to their history-based trust. Behavioral assessments of teammate interactions during task closure and reporting may be helpful to understanding enduring dispositional trust changes. Tuckman and Jensen's (1977) adjourning stage equates our Experience stage. Their adjourning stage addresses the end of the team life cycle, where members conclude team activity, assess overall team performance and dynamics, and return to a post-team solitary state.

In the team trust life cycle, the Experience stage influences both the Initial stage and the Orientation stage. Over time, team participation and experience working with autonomy can alter an individual's future propensity to trust regardless of future team inclusion. For teams that regroup for future tasking, learned trust affects their perceptions of known teammates and familiar autonomy, thus shortening subsequent Orientation stages.

Case Study Exemplar: The team arrives at their final destination or end goal. They assess their performance, the autonomy's performance, and complete after-action reports and reviews. These final interactions can be greatly influential to their long-term trust and recollections of team membership. This is the last opportunity to gather assessments of team trust, which would be in the form of subjective surveys and behavioral assessments. Individual trust assessments can also occur and be examined for changes in dispositional trust due to team membership.

Suggested Assessments and Methods: In the Experience stage, assessments for the human teammates consist largely of self-report surveys (both individual and team). Changes in the autonomy programming can be reviewed for changes after interacting with the team. Refer to the Toolkit section (Section 4) for detailed information about each assessment tool.

- Surveys
 - Integrated Model of Trust Scale (Muir and Moray 1996)
 - Human–Robot Interaction Trust Scale (Yagoda and Gillan 2012)
 - System Trustworthiness Scale (Schaefer et al. 2012)
 - Checklist for Trust in People and Automation (Jian et al. 2000)
 - Trust Perception Scale (Schaefer 2013)
- Behavioral measures
 - Proximity
 - Facial expressions, gestures, behavioral synchrony
- Communication measures, rate, flow, and content
- Physiological measures
 - None at this time
- Autonomy assessments
 - Post-task performance review
 - Algorithmic changes

5.2.1 Team Dynamics

As stated earlier, the team itself can impact how trust analyses should be interpreted. Trust is known to be context and task dependent, and as such the associated team dynamics can vary the descriptions and interpretation of the data. The size of the team adds to the complexity by muddying the waters and injecting uncertainty—did a person’s trust rating come from a specific interaction, or was it influenced by secondary or even tertiary trust from other team members (Schaefer et al. 2015)? Further, the technology, including the availability of different types and levels of autonomy as well as methods for communicating intent or reasoning, can further constrain the trust interpretation as interaction modality determines the cues you gather from others. Access to this information or interpretation of the team members’ capabilities is an important consideration when describing trust outcomes. Arguably, trust can be developed more easily with a team member or IA that you can physically see (embodied with local proximity) versus an embedded autonomy or a team member you can only “hear” [Baker et al. 2018]).

5.2.2 Individual Differences

A second major constraint to appropriately interpreting team trust analyses is individual differences. The impact of individual differences on trust development and calibration is well documented in the literature (see Schaefer et al. 2016). For example, the appropriate level of autonomy may be dependent on the individual characteristics of the human team members (Parasuraman et al. 2009; Chen et al. 2011). Highly skilled operators may become bored and complacent if the level of automation is too high, meaning that disengagement is not indicative of distrust for skilled operators, while lesser-skilled operators may depend on a higher level of automation to maintain acceptable performance and not become overwhelmed, meaning that engagement is indicative of trust for unskilled operators (Manzey et al. 2008; Wright et al. 2018). In addition, factors such as spatial ability, attentional control, and working memory capacity have shown to be differentiating factors for selecting the appropriate automation level (Szalma and Taylor 2011; Ahmed et al. 2014; Chen and Barnes 2014; Wright et al. 2018). Yet, when trying to identify the implications for team trust, this area becomes even more complex when trying to account for the global impact from each team member's individual differences. That said, balancing this constraint is critical for maintaining and calibrating trust. Prior research has shown that a combination of age, personality, and response to uncertainty in line with perceived stress, workload, and trust, can drastically affect the interpretation of self-reported trust and behavioral analyses (Neubauer et al. 2020).

5.2.3 Culture

When considering trust as a temporal construct, culture can directly affect the interpretation of the trust measures. In a highly mobile, individualistic culture, such as in the United States of America, early trust benefits the trustor (i.e., they tend to trust earlier in the relationship) because an individual's social relationships may be less established (Yamagishi et al. 1998; Macy and Sato 2002). However, in a collectivist culture with stable long-term social relationships, such as in Japan, individuals rely less upon generalized trust in establishing new relations, as these relations are initiated largely based upon recommendations from already-trusted associates. As such, the low-trust cultures tend to be less "committed" in early trust stages than the high-trust cultures (Yamagishi and Yamagishi 1994; Yuki et al. 2005). A consequence of this difference in high- versus low-trust culture is how the timing of trust violations impact the relationship. Early trust violations are much more damaging in a high-trust culture than in a low-trust culture, while late trust violations are more damaging in low-trust culture relationships (Kuwabara et al. 2014).

This issue is further complicated in the team trust measurement dynamic because literature suggests that the formation of interpersonal trust differs from that of human–automation trust (Madhavan and Wiegmann 2007). Specifically, interpersonal trust is initially founded on the predictability of another person’s actions because people tend to be rather cautious when entering relationships with strangers. Over time, as the relationship progresses, the dependability or integrity of the other party then becomes the basis of trust. Interpersonal relationships that are mature are based on faith or benevolence (Hoff and Bashir 2015). In contrast, the progression of trust in automation is often the opposite. That is, many people tend to carry an automation bias, also discussed as a positivity bias, that manifests in the presence of or during initial interactions with novel automated systems and these can lead to instances of complacency and, ultimately, misuse of the IA (Manzey et al. 2008; Parasuraman and Manzey 2010; Kok and Soh 2020). A common assumption that appears to underlie this bias has been recently discussed as a so-called Perfect Automation Schema, wherein people assume that machines operate, or are supposed to operate, without any level of error (Merritt et al. 2013). As a result, initial trust in these systems is based on faith—or more generally, expectations based on minimal prior observations—however, it is quickly challenged when system errors occur; as the relationship progresses, dependability and predictability govern the trust relationship (Hoff and Bashir 2015).

5.2.4 Attitudes toward Autonomy

The final consideration for appropriately interpreting team trust is constrained by attitudes toward autonomy. Automation tends to be viewed rigidly by operators—it works, or it does not (Ross et al. 2007). This polarization bias could have far-reaching consequences when the autonomy has multitasking capabilities (Beck et al. 2002). Highly reliable automation encourages operator complacency, or the “non-vigilance based on an unjustified assumption of satisfactory system state” (Parasuraman et al. 1993, p.2). This is particularly evident when the operator has other demands on their attention, such as in a multitasking situation (Chen and Barnes 2014). However, witnessing autonomy errors not only damages the operator’s trust in the autonomy, it also reduces the operator’s confidence in their own abilities to accurately assess the autonomy’s reliability (Wright et al. 2017; Wright et al. 2020). Understanding the autonomy’s underlying reasoning can mitigate the effects of unreliable autonomy, thus bolstering operator trust in the autonomy (Hoff and Bashir 2015; Wright et al. 2020).

5.3 Considerations for Team Trust Measurement

While multiple types of trust exist, trust is a multilevel concept as well. In a trust relationship, the trustor and referent of trust may be singular or multiple. That is, trust relations may exist between two individuals, between an individual and a group, between a group and an individual, and between groups. Which level(s) of trust one wishes to capture will shape which measures to use to capture trust. Many survey instruments effectively capture different levels of trust. For example, a survey from Schoorman et al. (2007) uses items where the referent and trustor are singular: “My supervisor keeps my interests in mind when making decisions” and “It is important for me to have a good way to keep an eye on my supervisor,” for example. Serva et al. (2005) use first-person plural items to capture a group of trustors: “Our development team felt that the management team was very capable of performing its job” and “Our development team believed that the management team tried to be fair in dealings with others,” for example. This approach allows more effective measurement of a group’s perception of trust in another rather than a specific individual’s trust relations. Capturing both individual and group-level perceptions of trust, Mayer and Davis (1999) use a mix of items: “Top management really looks out for what is important to me,” “Top management tries hard to be fair in dealings with others,” and “If I had my way, I wouldn’t let top management have any influence over issues that are important to me.” Just like in human–human teams, trust in human–autonomy teams has multiple factors and may occur at multiple levels and the factors and levels must align with the instruments used to measure trust. Indeed, some recent surveys of trust in human–autonomy teams have accounted for multiple factors and levels of trust. Yagoda and Gillan’s (2012) Human Robot Interaction Trust Scale, for example, captures trust in one’s own skills, knowledge of tasks and systems, and familiarity with one’s teammates; trust in the team’s planning, leadership, communication, and decision-making; and trust in different components of the autonomous agent such as navigation capabilities and level of automation. When measuring trust within a team, whether that trust spans individuals, groups, or both, aligning the levels of the trustors and referents of trust is essential for achieving a valid trust measure.

Outside these survey instruments, capturing plural trustors or referents of trust becomes more challenging. Quantification of trust relations may be achieved through several behavioral measures, such as number or length of instances of attentional monitoring and number of instances of reliance or compliance. Measured at the individual level, these two measures require careful consideration when calculating trust where the trustor or referent of trust spans more than one individual. Team-level measures of trust may be aggregated across the behaviors of all members of the team, but how to aggregate behaviors is not always

straightforward. One may choose to aggregate through one of several approaches: mean, additive, conjunctive, or disjunctive. As behavioral measures increasingly complement survey-based measures of trust, topics of aggregation deserve additional attention to maximize the validity of those behavioral measures compared to more traditional survey measures that more easily capture multiple levels of trust.

Some behavioral measures are inherently relational and more amenable to scaling from individuals to groups. Proxemics, entrainment, and facial-expression matching are all relational measures based on multiple individuals and can be trivially scaled from dyads to larger groups. While all three measures capture differences between individuals, these can be scaled from differences between dyads to differences across triads, tetrads, or larger groups. However, some consideration must be given to how to characterize the differences in groups—for example, determining if the overall group-level difference should be equal to the average difference, the largest difference, or the smallest difference. But as inherently relational measures, these more easily capture inherently relational team states like trust.

One final challenge for capturing team trust at multiple levels is the limitations on inferring autonomous agents' trust in humans (Kok and Soh 2020). Autonomous agents cannot (yet?) reliably fill out survey measures, and physiological measures like pupillometry, HRV, and EDA do not apply to autonomous agents. Rather, trust measures will depend largely on what sort of information may be gleaned from autonomous agents, such as data logs. Behavioral measures may be key here, as they may be most analogous to human behaviors used to measure trust. Some such behaviors could include an autonomous vehicle's decision to override a human's lane change while another vehicle occupies the blind spot, or suggesting an alternative course of action to a human whose plan has a low probability of achieving group goals. As humans and autonomous agents increasingly begin to work together on teams, these sorts of behavioral measures for inferring autonomous agents' trust in humans deserve considerable attention.

6. Conclusions and Path Forward

Measuring and interpreting trust between dyads is incredibly complex and is an area of ongoing research by the trust community. This complexity is further exacerbated when expanding beyond this dyadic relationship toward a larger team trust, whereby a team includes multiple people and multiple IAs (either embedded or embodied). As we continue to explore team trust dynamics more fully, this report serves first as a critical steppingstone to assist researchers in effectively selecting

appropriate measures given the temporal nature of trust and growing complexity of teams operating in complex, dynamic environments. While this toolkit provides a novel approach for assessing team trust, it further addresses the complexities of interpreting trust scores by providing a foundational model for assessing the trust of not just an individual, but a team. Assessing trust is a temporal process that occurs throughout the entire life cycle of a team and builds on four basic steps: the *initial trust* stage to assess trust propensity; the *orientation stage* to set an individual's expectations for the team and upcoming task; the *trust calibration stage* to build trust for working with a specific team during and after an interaction; and the *experience stage*, which highlights the trust dynamics that become more, or less, important over time through this continuous, cyclical process of trust development, maintenance, and calibration, until a team disbands. These steps integrate the six types of trust that have been shown to be important for team trust measurement: trust propensity, affect-based trust, cognitive or history-based trust, trustworthiness, situational trust, and learned trust. Finally, effective interpretation of team trust is constrained by several variables including team dynamics, individual differences, culture, and attitudes toward autonomy.

Overall, measurement of trust in human–autonomy teams remains a complex problem and, as a result, there is no one-size-fits-all approach. Rather, researchers must continually assess which types of measures are best suited to the context, recognizing that different measures will have a stronger impact on appropriately quantifying trust at a given time. Therefore, for many applications, a multimethod approach is warranted (see also Schaefer et al. 2019b; Milner et al. 2020; Schaefer et al. 2021). Continued research building on this toolkit will support the development of more appropriate metrics of team trust that will help us understand how human–autonomy teams perform, especially as autonomous capabilities increase, and when interventions are needed, providing a clearer picture of how the team is functioning, interacting, and accomplishing goals.

Path Forward

To shed light on the path forward, we highlight three main areas of research: further validation of trust measures, synthesizing measures into metrics, and specifying how trust measurement should inform appropriate trust-based interventions that facilitate more effective team performance. The first step is to further refine and validate some of the exploratory measures listed in this toolkit. Specifically, communication flow, communication rates, physiological features, entrainment, and facial-analysis methods all show promise as direct or proxy measures of trust, but they have yet to be fully validated. The second step is to build on this measurement research to quantify appropriate metrics of team trust. While the measures provide values for specific constructs, metrics will provide a more

standardized assessment strategy for evaluating team trust in the human autonomy teaming context; in so doing, we will further support our multimethod, multimodal approach by leveraging the relationships that currently exist between these aforementioned trust measures, enabling us to more accurately understand trust development, maintenance, and calibration. Ultimately, understanding the causal structure that underlies the relationships between the underlying constructs of team trust and their measures will inform effective strategies for trust-based interventions.

Finally, appropriately measuring trust only gets us so far. The third step is determining how to effectively use that measurement to provide critical and appropriate trust-based interventions to improve team performance. Specifically, these interventions can be directed toward several possible changes in the team operations including training recommendations, changes in autonomy behavior, implementation of algorithmic assurances in the autonomy, improving communication and transparency elements, or even supporting after-action reviews. To that end, our research is being used to inform the development of two technologies: a software toolkit and a global after-action review technology that will help us assess team interactions in near real-time, and through algorithm creation and visualization techniques we will be able to observe changes in trust over time and identify areas where an intervention is warranted. Although still early in the development process, these technologies, coupled with the research presented here, will enable researchers to develop more precise, valid measures of trust and deploy effective, trust-enhancing interventions in practical settings.

7. References

- Abrams LC, Cross R, Lesser E, Levin DZ. Nurturing interpersonal trust in knowledge-sharing networks. *Academy of Management Perspectives*. 2003;17(4):64–77.
- Ahmed N, de Visser E, Shaw T, Mohamed-Ameen A, Campbell M, Parasuraman, R. Statistical modelling of networked human–automation performance using working memory capacity. *Ergonomics*. 2014. doi:0.1080/00140139.2013.855823.
- Amemiya S, Ohtomo K. Effect of the observed pupil size on the amygdala of the beholders. *Soc Cogn Affect Neurosci*. 2012;7:332–341. doi:10.1093/scan/nsr013.
- Astolfi L, Toppi J, De Vico Fallani F, Vecchiato G, Salinari S, Mattia D, Cincotti F, Babiloni F. Neuroelectrical hyperscanning measures simultaneous brain activity in humans. *Brain Topography*. 2010;23(3):243–256.
- Atkinson D, Hancock P, Hoffman RR, Lee JD, Rovira E, Stokes C, Wagner AR. Trust in computers and robots: the uses and boundaries of the analogy to interpersonal trust. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*; 2012 Oct 22–26; Boston, MA. HFES 2012.p. 303–307. doi.org/10.1177/1071181312561071.
- Bagheri N, Jamieson GA. The impact of context-related reliability on automation failure detection and scanning behaviour. In: 2004 IEEE International Conference on Systems, Man and Cybernetics; 2004 Oct. IEEE. Vol. 1; p. 212–217. IEEE Cat. No. 04CH37583.
- Baker AL, Phillips EK, Ullman D, Keebler JR. Toward an understanding of trust repair in human–robot interaction: current research and future directions. *ACM Transactions on Interactive Intelligent Systems*. 2018;8(4)30:1–30.
- Baker AL, Fitzhugh SM, Forster DE, Brewer RW, Krausman A, Scharine A, Schaefer KE. Team trust in human–autonomy teams: analysis of crew communication during manned–unmanned gunnery operations. DEVCOM Army Research Laboratory; 2020. Report No.: ARL-TR-8969.
- Baker AL, Fitzhugh SM, Huang L, Forster DE, Scharine A, Neubauer C, Lematta G, Bhatti S, Johnson CJ, Krausman A, Holder E, Schaefer KE, Cooke NJ. Approaches for assessing communication in human–autonomy teams. *Human–Intelligent Systems Integration*. 2021;3(2):99–128. doi.org/10.1007/s42454-021-00026-2.

- Balliet D, Wu Y, De Dreu CKW. In-group favoritism and cooperation: a meta-analysis. *Psychological Bulletin*. 2014;140:1556–81.
- Barrick MR, Stewart GL, Neubert MJ, Mount MK. Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*. 1998;83:377–391.
- Bartlett B, Estivill-Castro V, Seymon S. Dogs or robots: why do children see them as robotic pets rather than canine machines? *Proceedings of the Fifth Conference on Australasian User Interface*; 2004. Vol. 28; p. 7–14.
- Beck HP, Dzindolet MT, Pierce LG. Operators' automation usage decisions and the sources of misuse and disuse. *Advances in Human Performance and Cognitive Engineering Research*. 2002;2:37–78.
- Behrens F, Snijdwint JA, Moulder RG, Prochazkova E, Sjak-Shie EE, Boker SM, Kret ME. Physiological synchrony promotes cooperative success in real-life interactions. *bioRxiv*. 2019;792416.
- Bernieri FJ, Rodenthal R. Interpersonal coordination: behavior matching and interaction synchrony. In: Feldman RS, Rime B, editors. *Fundamentals of Nonverbal Behavior: Studies in Emotion and Social Interaction*. Cambridge University Press; 1991. p. 401–432.
- Bethel JL, Murphy RR, Salomon K. Psychophysiological experimental design for use in human–robot interaction studies. In: *International Symposium on Collaborative Technologies and Systems*; 2007. CTS 2007.
- Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of Machine Learning Research*. 2003;3(Jan):993–1022.
- Bonebright D. 40 years of storming: a historical review of Tuckman's model of small group development. *Human Resource Development International*; 2010;13:111–120. doi:10.1080/13678861003589099.
- Borum R. The science of interpersonal trust. *Mental Health Law and Policy Faculty Publications*. University of South Florida; 2010. Faculty Publications No.: 574. http://scholarcommons.usf.edu/mhlp_facpub/574.
- Bradshaw J. Pupil size as a measure of arousal during information processing. *Nature*. 1967;216:515–516.
- Breazeal C. Regulation and entrainment in human–robot interaction. *The International Journal of Robotics Research*. 2002;21(10–11):883–902.

- Bromiley P, Cummings LL. Transaction costs in organizations with trust. In: Bies R, Sheppard B, Lewicki R, editors. *Research on Negotiations in Organizations*; vol. 5; 1995.
- Burke C, Sims DE, Lazzara EH, Salas E. Trust in leadership: a multi-level review and integration. *Leadership Quarterly*. 2007;18:606–632. doi:10.1016/j.leaqua.2007.09.006.
- Burt RS, Knez M. Kinds of third-party effects on trust. *Rationality and Society*. 1995;7(3):255–292. doi:10.1177/1043463195007003003.
- Cannon-Bowers JA, Salas E. Reflections on shared cognition. *Journal of Organizational Behavior*. 2001;22(2):195–202.
- Carmody PC, Mateo JC, Bowers D, McCloskey M. Linguistic coordination as an unobtrusive, dynamic indicator of rapport, prosocial team processes, and performance in team communication. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2017 Sep. SAGE Publications; 2017;61(1):140–144.
- Catenaccio P. Markers of trust: epistemic adverbs of certainty and restrictive adverbs in CSR reports. *Linguistics*; 2017.
- Ceron A. Internet, news, and political trust: the difference between social media and online media outlets. *Journal of Computer-Mediated Communication*. 2015;20(5):487–503.
- Chancey ET, Bliss JP, Yamani Y, Handley H. Trust and the compliance-reliance paradigm: the effects of risk, error bias, and reliability on trust and dependence. *Human Factors*. 2017;59(3):333–345.
- Chartrand TL, Bargh JA. The chameleon effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology*. 1999;76:893–910.
- Chen JYC, Barnes MJ. Supervisory control of multiple robots: effects of imperfect automation and individual differences. *Human Factors*. 2012;54(2):157–174.
- Chen JYC, Barnes MJ. Human–agent teaming for multirobot control: a review of human factors issues. *IEEE Trans Hum–Mach Sys*. 2014;44(1):13–29.
- Chen JY, Chen, JYC, Barnes, MJ, Harper-Sciarini M. Supervisory control of multiple robots: human performance issues and user-interface design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 2011;41(4):435–454.

- Chen JYC, Lakhmani SG, Stowers K, Selkowitz AR, Wright JL, Barnes M. Situation awareness-based agent transparency and human–autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science*. 2018;19(3):259–282.
- Chen S, Andrienko N, Andrienko G, Adilova L, Barlet J, Kindermann J, Nguyen PH, Thonnard O, Turkay C. LDA ensembles for interactive exploration and categorization of behaviors. *IEEE Transactions on Visualization and Computer Graphics*; 2019;26(9):2775–2792.
- Chu VC, Lucas GM, Lei S, Mozgai S, Khooshabeh P, Gratch J. Emotion regulation in the prisoner’s dilemma: effects of reappraisal on behavioral measures and cardiovascular measures of challenge and threat. *Front Hum Neurosci*. 2019;13:50.
- Cikara M, Bruneau E, Van Bavel JJ, Saxe R. Their pain gives us pleasure: how intergroup dynamics shape empathic failures and counter-empathic responses. *Journal of Experimental Social Psychology*. 2014;55:110–25. doi:10.1016/j.jesp.2014.06.007.
- Clark HH, Brennan SE. Grounding in communication. In: Resnick LB, Levine JM, Teasley SD, editors. *Perspectives on socially Shared Cognition*. APA Books; 1991. p. 127–149.
- Colquitt J, Scott B, LePine J. Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *The Journal of Applied Psychology*. 2007;92:909–27. doi:10.1037/0021-9010.92.4.909.
- Costa AC, Anderson N. Measuring trust in teams: development and validation of a multifaceted measure of formative and reflective indicators of team trust. *European Journal of Work and Organizational Psychology*. 2011;20(1):119–154. doi:10.1080/13594320903272083.
- Costa AC, Anderson N. Team trust. In: Salas E, Rico R, Passmore J, editors. *The Wiley Blackwell handbook of the psychology of team working and collaborative processes*. John Wiley & Sons Ltd; c2017. p. 393–416. doi:10.1002/9781118909997.
- Costa AC, Fulmer CA, Anderson N. Trust in work teams: an integrative review, multilevel model, and future directions. *Journal of Organizational Behavior*. 2018;39(2):169–184. doi:org/10.1002/job.2213.
- Cramton CD. The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science*. 2001;12(3):346–371.

- Dale R, Kirkham NZ, Richardson DC. The dynamics of reference and shared visual attention. *Frontiers in Psychology*. 2011;2:355.
- Dautenhahn K. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B*. 2007;362:679–04.
- De Jong BA, Elfring T. How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Academy of Management Journal*. 2010;53(3):535–549.
- De Jong BA, Dirks KT. Beyond shared perceptions of trust and monitoring in teams: implications of asymmetry and dissensus. *Journal of Applied Psychology*. 2012;97(2):391–406.
- De Jong BA, Dirks KT, Gillespie N. Trust and team performance: a meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*. 2016;101(8):1134–1150.
- De Looze C, Scherer S, Vaughan B, Campbell N. Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication*. 2014;58:11–34.
- de Visser EJ, Cohen M, Freedy A, Parasuraman R. A design methodology for trust cue calibration in cognitive agents. *International Conference on Virtual, Augmented and Mixed Reality*; 2014. Springer, Cham. p. 251–262.
- de Visser EJ, Peeters MM, Jung MF, Kohn S, Shaw TH, Pak R, Neerincx MA. Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics*. 2020;12:459–478.
- DeChurch LA, Mesmer-Magnus JR. The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of Applied Psychology*. 2010;95(1):32–53.
- Deerwester S, Dumais ST, Furnas GW, Landauer TK, Harshman R. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*. 1990;41(6):391–407.
- Delgado-Márquez BL, Hurtado-Torres NE, Aragón-Correa JA. The dynamic nature of trust transfer: measurement and the influence of reciprocity. *Decision Support Systems*. 2012;54:226–234. 10.1016/J.DSS.2012.05.008.
- Deligianis C, Stanton CJ, McGarty C, Stevens CJ. The impact of intergroup bias on trust and approach behaviour towards a humanoid robot. *Journal of Human–Robot Interaction*. 2017;6(3):4–20.

- Demir M, Likens AD, Cooke NJ, Amazeen PG, McNeese NJ. Team coordination and effectiveness in human–autonomy teaming. *IEEE Transactions on Human–Machine Systems*. 2019;49(2):150–159. doi:10.1109/THMS.2018.2877482.
- Demos KE, Kelley WM, Ryan SL, Davis FC, Whalen PJ. Human amygdala sensitivity to the pupil size of others. *Cereb Cortex*. 2008;18:2729–2734.
- Desai M, Kaniarasu P, Medvedev M, Steinfeld A, Yanco H. Impact of robot failures and feedback on real-time trust. Paper presented at HRI '13: the 8th ACM/IEEE International Conference on Human–Robot Interaction; 2013 Mar 3–6; Tokyo, Japan.
- Deutsch M. Trust and suspicion. *Journal of Conflict Resolution*. 1958;2(4):265–279. doi:10.1177/002200275800200401.
- Dimberg U. Facial reactions to facial expressions. *Psychophysiology*. 1982;19:643–647.
- Dixon SR, Wickens CD. Automation reliability in unmanned aerial vehicle control: a reliance-compliance model of automation dependence in high workload. *Human Factors*. 2006;48(3):474–486.
- Dixon S, Wickens C, McCarley J. On the independence of compliance and reliance: are automation false alarms worse than misses? *Human Factors*. 2007;49:564–72. doi:10.1518/001872007X215656.
- Donohue W, Liang Y, Druckman D. Validating LIWC dictionaries. *Journal of Language and Social Psychology*. 2014;33:282–301.
- Dowell N, Nixon T, Graesser A. Group communication analysis: a computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*. 2019;51:1007–1041.
- Drnec K, Metcalfe J. Paradigm development for identifying and validating indicators of trust in automation in the operational environment of human automation integration. In: Schmorrow D, Fidopiastis C, editors. *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*. AC 2016. Lecture Notes in Computer Science vol 9744. Springer, Cham. doi.org/10.1007/978-3-319-39952-2_16.
- Drnec K, Paul VJ, Haynes B, Christensen JC, Lukos JR, Marathe AR, Metcalfe JS. Trust in driving automation: reliance does not always imply trust. *Frontiers in Robotics and AI*. To be published.

- Eckmann J-P, Oliffson Kamphorst S, Ruelle D. Recurrence plots of dynamical systems. *Europhysics Letters*. 1987;4:973–977.
- Ekman P. Universal facial expressions of emotion. *California Mental Health Research Digest*. 1970;8(4):151–158.
- Ekman P, Friesen WV. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*. 1976;1:56–75.
- Elkins AN, Muth ER, Hoover AW, Walker AD, Carpenter TL, Switzer FS. Physiological compliance and team performance. *Applied Ergonomics*. 2009;40: 997–1003.
- Endsley MR. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 1995;37(1):32–64.
- Evans AM, Revelle W. Survey and behavioral measurements of interpersonal trust. *Journal of Research in Personality*. 2008;42(6):1585–1593. doi.org/10.1016/j.jrp.2008.07.011.
- Farroni T, Csibra G, Simion F, Johnson MH. Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences (PNAS)*. 2002;99:9602–9605.
- Feitosa J, Grossman R, Kramer WS, Salas E. Measuring team trust: a critical and meta-analytical review. *Journal of Organizational Behavior*. 2020;41(5):479–501.
- Ferrin DL, Dirks KT, Shah PP. Direct and indirect effects of third-party relationships on interpersonal trust. *Journal of Applied Psychology*. 2006;91(4):870–883. doi:10.1037/0021-9010.91.4.870.
- Fischer U, McDonnell L, Orasanu J. Linguistic correlates of team performance: toward a tool for monitoring team functioning during space missions. *Aviation, Space, and Environmental Medicine*. 2007;78(5):B86–B95.
- Freedly A, de Visser E, Weltman G, Coeyman N. Measurement of trust in human–robot collaboration. *Proceedings of 2007 International Symposium on Collaborative Technologies and Systems*; 2007 May 25; Orlando, FL. p. 106–114. doi:10.1109/CTS.2007.4621745.
- Frimer JA, Schaefer NK, Oakes H. Moral actor, selfish agent. *Journal of Personality and Social Psychology*. 2014;106(5):790.

- Fulmer CA, Gelfand MJ. At what level (and in whom) we trust: trust across multiple organizational levels. *Journal of Management*. 2012;38(4):1167–1230. doi:10.1177/0149206312439327.
- Gagge AP, Stolwijk JAJ, Hardy JD. Comfort and thermal sensations and associated physiological responses at various ambient temperatures. *Environmental Research*. 1967;1(1):1–20.
- Gefen D, Fresneda JE, Larsen KR. Trust and distrust as artifacts of language: a latent semantic approach to studying their linguistic correlates. *Frontiers in Psychology*. 2020;11:561. doi.org/10.3389/fpsyg.2020.00561.
- Geiskkovitch DY, Thiessen R, Young JE, Glenwright MR. What? That's not a chair!: how robot informational errors affect children's trust towards robots. In: 2019 14th ACM/IEEE International Conference on Human–Robot Interaction (HRI); 2019 Mar. IEEE. p. 48–6.
- Giles H. Communication accommodation theory. *The International Encyclopedia of Communication Theory and Philosophy*. 2016;1–7.
- Gill H, Boies K, Finegan JE, McNally J. Antecedents of trust: establishing a boundary condition for the relation between propensity to trust and intention to trust. *Journal of Business Psychology*. 2005;19(3):287–302.
- Glaholt MG, Reingold EM. Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*. 2011;4.
- Gonzalez AL, Hancock JT, Pennebaker JW. Language style matching as a predictor of social dynamics in small groups. *Communication Research*. 2010;37(1):3–19.
- Gorman JC, Cooke NJ, Amazeen PG, Fouse S. Measuring patterns in team interaction sequences using a discrete recurrence approach. *Human Factors*. 2012;54(4):503–17.
- Groom V, Nass C. Can robots be teammates? Benchmarks in human–robot teams. *Interaction Studies*. 2007;8(3):483–500.
- Grossman R, Feitosa J. Team trust over time: modeling reciprocal and contextual influences in action teams. *Hum Resour Manag Rev*. 2018;28:395–410.
- Gutzwiller RS, Chiou EK, Craig SD, Lewis CM, Lematta GJ, Hsiung C-P. Positive bias in the 'Trust in Automation Systems Survey'? An examination of the Jian et al. (2000) scale. *Proceedings of the Human Factors and Ergonomics Society*; 2019;63(1):217–221.

- Haring KS, Matsumoto Y, Watanabe K. How do people perceive and trust a lifelike robot. *Proceedings of the World Congress on Engineering and Computer Science*. 2013;1.
- Harrison NA, Singer T, Rotshtein P, Dolan RJ, Critchley HD. Pupillary contagion: central mechanisms engaged in sadness processing. *Social Cognitive and Affective Neuroscience*. 2006;1:5–17.
- Harrison NA, Gray MA, Critchley HD. Dynamic pupillary exchange engages brain regions encoding social salience. *Social Neuroscience*. 2009;4:233–243.
- Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Proceedings of Advances in Psychology*. 1988;52:139–183.
- Henkel Z, Bethel CL, Murphy RR, Srinivasan V. Evaluation of proxemic scaling functions for social robotics. *IEEE Transactions on Human–Machine Systems*. 2014;44(3):374–385.
- Henning RA, Boucsein W, Gil MC. Social-physiological compliance as a determinant of team performance. *International Journal of Psychophysiology*. 2001;40:221–232.
- Henning RA, Korbela K. Social-psychophysiological compliance as a predictor of future team performance. *Psychologia: An International Journal of Psychology in the Orient*. 2005;48:84–92.
- Henning RA, Armstead AG, Ferris JK. Social psychophysiological compliance in a four-person research team. *Applied Ergonomics*. 2009;40:1004–1010.
- Henshel D, Cains MG, Hoffman B, Kelley T. Trust as a human factor in holistic cyber security risk assessment. *Procedia Manufacturing*. 2015;3:1117–1124. doi:10.1016/j.promfg.2015.07.186.
- Hergeth S, Lorenz L, Vilimek R, Krems JF. Keep your scanners peeled: gaze behavior as a measure of automation trust during highly automated driving. *Human Factors*. 2016;58(3):509–519.
- Hess EH. The role of pupil size in communication. *Scientific American*. 1975;233(5):110–112,116–119.
- Hess U, Fischer A. Emotional mimicry as social regulation. *Personality and Social Psychology Review*. 2013;17:142–157.

- Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*. 2015;57(3):407–434. doi:10.1177/0018720814547570.
- Hung H, Gatica-Perez D. Estimating cohesion in small groups using audio–visual nonverbal behavior. *IEEE Transactions on Multimedia*. 2010;12(6):563–575. doi:10.1109/TMM.2010.2055233.
- Iio T, Shiomi M, Shinozawa K, Miyashita T, Akimoto T, Hagita N. Lexical entrainment in human–robot interaction: can robots entrain human vocabulary? 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems; 2009 Oct. IEEE. p. 3727–3734.
- Isberner MB, Richter T, Kaakinen J. Epistemic modality in sentence comprehension: effects of epistemic adverbs on eye movements. *The Role of Epistemic Monitoring in Language Comprehension*. 2013;137.
- Israelsen BW, Ahmed NR. “Dave... I can assure you... that it’s going to be all right...” A definition, case for, and survey of algorithmic assurances in human–autonomy trust relationships. *ACM Computing Surveys (CSUR)*. 2019;51(6):1–37.
- Izard CE. Forms and functions of emotions: matters of emotion–cognition interactions. *Emotion Review*. 2011;3(4):371–378. doi:10.1177/1754073911410737.
- Jackson KF, Prasov Z, Vincent EC, Jones EM. Draper trust assessment framework – trust assessment scales (v1.0). Draper; 2016.
- Jarvenpaa S, Knoll K, Leidner D. Is anybody out there? Antecedents of trust in global teams. *Journal of Management Information Systems*. 1998;14:29–64. doi:10.1080/07421222.1998.11518185.
- Jessup SA. Measurement of the propensity to trust automation [master’s thesis]. Wright State University; 2018.
- Jian JY, Bisanz A, Drury CG. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*. 2000;4:53–71. doi:10.1207/S15327566IJCE0401_04.
- Jøsang A, Presti SL. Analysing the relationship between risk and trust. In: Jensen C, Poslad S, Dimitrakos T, editors. *Trust Management. iTrust 2004*. Springer, Berlin, Heidelberg; 2004. *Lecture Notes in Computer Science*; vol. 2995. doi.org/10.1007/978-3-540-24747-0_11.

- Judson J. Dirty work: robots take on complex obstacles in US Army exercise. *Defense News*. 2019 May. <https://www.defensenews.com/land/2019/05/10/dirty-work-robots-take-on-complex-breach-as-army-evaluates-potential/>.
- Just MA, Carpenter PA. Eye fixations and cognitive processes. *Cognitive Psychology*. 1976;8(4):441–480.
- Kelsey M, Palumbo RV, Urbaneja A, Akcakaya M, Huang J, Kleckner IR, Barrett LF, Quigley KS, Sejdic E, Goodwin MS. Artifact detection in electrodermal activity using sparse recovery. In: Ahmad F, editor. *Proceedings of Compressive Sensing VI: From Diverse Modalities to Big Data Analytics*; 2017 May; Anaheim, CA. International Society for Optics and Photonics. Vol. 10211, p. 102110D.
- Keltner D, Ekman P, Gonzaga GC, Beer J. Facial expression of emotion. In: Davidson RJ, Scherer KR, Goldsmith HH, editors. *Handbook of Affective Sciences*. Oxford University Press; 2003. Series in Affective Science. p. 415–432.
- Kelton K, Fleischman KR, Wallace WA. Trust in digital information. *Journal of the American Society for Information Science and Technology*. 2007;59(3):363–374. doi.org/10.1002/asi.20722.
- Kim G, Koo H. The causal relationship between risk and trust in the online marketplace: a bidirectional perspective. *Computers in Human Behavior*. 2016;55:1020–1029. doi.org/10.1016/j.chb.2015.11.005.
- Kim PH, Ferrin DL, Cooper DL, Dirks KT. Removing the shadow of suspicion: the effects of apology versus denial for repairing competence versus integrity-based trust violations. *J Appl Psychol*. 2004;89(1):104–118.
- Kindervag J. No more chewy centers: introducing the zero trust model of information security. *Forester Research*; 2010.
- Kohn SC, de Visser EJ, Wiese E, Lee Y-C, Shaw TH. Measurement of trust in automation: a narrative review and reference guide. *Front Psychol*. 2021;12:604977. doi:10.3389/fpsyg.2021.604977.
- Kok BC, Soh H. Trust in robots: challenges and opportunities. *Curr Robot Rep*. 2020;1(4):297–309. doi: 10.1007/s43154-020-00029-y. Epub 2020 Sep 3. PMID: 34977590; PMCID: PMC7467858.

- Kollock P. The emergence of exchange structures: an experimental study of uncertainty, commitment, and trust. *American Journal of Sociology*. 1994;100(2):313–345. doi:10.1086/230539.
- Kondo N, Tominaga H, Shibasaki M, Aoki K, Okada S, Nishiyasu T. Effects of exercise intensity on the sweating response to a sustained static exercise. *Journal of Applied Physiology*. 2000;88(5):1590–1596.
- Konvalinka I, Xygalatas D, Bulbulia J, Schjodt U, Jegindo E-M, Wallot S, Van Orden G, Roepstorff A. Synchronized arousal between performers and related spectators in a fire-walking ritual. *PNAS. Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(20):8514–8519.
- Kramer RM. Trust and distrust in organizations: emerging perspectives, enduring questions. *Annual Review of Psychology*. 1999;50(1):569–598.
- Kret ME. Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Front Psychol*. 2015; 6:71.
- Kuwabara K, Vogt S, Watabe M, Komiya A. Trust, cohesion, and cooperation after early versus late trust violations in two-person exchange: the role of generalized trust in the United States and Japan. *Social Psychology Quarterly*. 2014;77(4):344–360.
- Lahijanian M, Kwiatkowska M. Social trust: a major challenge for the future of autonomous systems. 2016 AAAI Fall Symposium Series; 2016 Sep.
- Lakin JL, Chartrand TL. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science*. 2003;14:334–339.
- Lam C. Linguistic politeness in student-team emails: its impact on trust between leaders and members. *IEEE Transactions on Professional Communication*. 2011;54(4):360–375.
- Langer A, Fiengold-Polak R, Mueller O, Kellmeyer P, Levy-Tzedek S. Trust in socially assistive robots: considerations for use in rehabilitation. *Neuroscience & Biobehavioral Reviews*. 2019;104:231–239.
- Langfred C. Too much of a good thing? Negative effects of high trust and individual autonomy in self-managed teams. *Academy of Management Journal*. 2004;47:385–399. doi:10.2307/20159588.
- Lavín C, San Martín R, Jubal ER. Pupil dilation signals uncertainty and surprise in a learning gambling task. *Front Behav Neurosci*. 2014;7:218.

- Lawler KA, Younger JW, Piferi RL, Billington E, Jobe R, Edmondson K, Jones WH. A change of heart: cardiovascular correlates of forgiveness in response to interpersonal conflict. *Journal of Behavioral Medicine*. 2003;26(5):373–393.
- Lee JD, Moray N. Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human–Computer Studies*. 1994;40(1):153–184.
- Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Human Factors*. 2004;46(1):50–80.
- Levenson RW, Gottman JM. Marital interaction: physiological linkage and affective exchange. *Journal of Personality and Social Psychology*. 1983;45:587–597.
- Lewicki RJ. Trust, trust development, and trust repair. In: *The Handbook of Conflict Resolution: Theory and Practice*. Jossey-Bass;2007.
- Lewicki RJ, Tomlinson EC, Gillespie N. Models of interpersonal trust development: theoretical approaches, empirical evidence, and future directions. *Journal of Management*. 2006;32(6):991–1022.
- Lewis M, Sycara K, Walker P. The role of trust in human–robot interaction. In: Abbass H, Scholz J, Reid D, editors. *Foundations of Trusted Autonomy*. Springer-Verlag, Berlin-Heidelberg;2017. p. 135–159.
- Lewis M, Li H, Sycara K. Deep learning, transparency, and trust in human robot teamwork. In: *Trust in Human–Robot Interaction*. Academic Press; 2021. p. 321–352.
- Linde C. The quantitative study of communicative success: politeness and accidents in aviation discourse. *Language in Society*. 1988;375–399.
- Lindenberger U, Li SC, Gruber W, Muller V. Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neuroscience*. 2009;10:22. doi:10.1186/1471-2202-10-22.
- Lindskold S. Trust development, the GRIT proposal, and the effects of conciliatory acts on conflict and cooperation. *Psychological Bulletin*. 1978;85(4):772–793. doi.org/10.1037/0033-2909.85.4.772.
- Lu Y, Sarter N. Eye tracking: a process-oriented method for inferring trust in automation as a function of priming and system reliability. *IEEE Transactions on Human–Machine Systems*. 2019;49(6):560–568.
- Lu Z, Peng Y, Ip HH. Image categorization via robust pLSA. *Pattern Recognition Letters*. 2010;31(1):36–43.

- Lyons JB. Being transparent about transparency: a model for human–robot interaction. AAAI Spring Symposium: Trust and Autonomous Systems; 2013.
- MacArthur KR, Stowers K, Hancock PA. Human–robot interaction: proximity and speed—slowly back away from the robot! In: *Advances in Human Factors in Robots and Unmanned Systems*. Springer, Cham; 2017. p. 365–374.
- MacMillan J, Entin EE, Serfaty D. Communication overhead: the hidden cost of team cognition. Team cognition: process and performance at the inter- and intra-individual level. American Psychological Association; 2004. http://www.apptima.com/publications/2004_MacMillan_EntinEE_Serfaty.pdf.
- Macy MW, Sato Y. Trust, cooperation, and market formation in the US and Japan. *Proceedings of the National Academy of Sciences*. 2002;99(suppl 3):7214–7220.
- Madhavan P, Wiegmann DA. Similarities and differences between human–human and human–automation trust: an integrative review. *Theoretical Issues in Ergonomics Science*. 2007;8(4):277–301.
- Manser T, Howard SK, Gaba DM. Adaptive coordination in cardiac anaesthesia: a study of situational changes in coordination patterns using a new observation system. *Ergonomics*. 2008;51(8):1153–1178.
- Manzey D, Reichenbach J, Onnasch L. Performance consequences of automated aids in supervisory control: the impact of function allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2008;52(4):297–301.
- Marathe AR, Schaefer KE, Evans AW, Metcalfe JS. Bidirectional communication for effective human–agent teaming. In: Chen J, Fragomeni G, editors. *Virtual, Augmented, and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation*. VAMR 2018. Lecture Notes in Computer Science. 2018;10909. doi:10.1007/978-3-319-91581-4_25.
- Marsh KL, Richardson MJ, Baron RM. Contrasting approaches to perceiving and acting with others. *Ecological Psychology*. 2006;18:1–38.
- Marshall SP. Identifying cognitive state from eye metrics. *Aviat Space Environ Med*. 2007;78:B165–B175.
- Martin M, Foltz P. Automated team discourse annotation and performance prediction using LSA. *Proceedings of HLT-NAACL 2004: Short Papers*; 2004. p. 97–100.

- Mathôt S. Pupillometry: psychology, physiology, and function. *Journal of Cognition*. 2018;1(1).
- Matthews R, McDonald NJ, Trejo LJ. Psycho-physiological sensor techniques: an overview. In: Schmorrow DD, editor. *Foundations of Augmented Cognition*. Erlbaum;2005. p. 263–272.
- Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. *The Academy of Management Review*. 1995;20(3):709–734.
- Mayer RC, Davis JH. The effect of the performance appraisal system on trust for management: a field quasi-experiment. *Journal of Applied Psychology*. 1999;84(1):123.
- McAllister D. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*. 1995;38:24–59. doi:10.2307/256727.
- McKnight DH, Cummings LL, Chervany NL. Initial trust formation in new organizational relationships. *Academy of Management Review*. 1998;23(3):473–490.
- McNeese N, Demir M, Chiou E, Cooke N, Yanikian G. Understanding the role of trust in human–autonomy teaming. Paper presented at the 52nd Hawaii International Conference on System Sciences; 2019; Grand Wailea, Maui.
- Mehler B, Reimer B, Coughlin JF, Dusek JA. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*. 2009;2138:6–12. Transportation Research Board of the National Academies. doi:10.3141/2138-02.
- Mercado JE, Rupp MA, Chen JY, Barber D, Procci K, Barnes M. Effects of agent transparency on multi-robot management effectiveness. *Army Research Laboratory (US)*; 2015. Report No.: ARL-TR-7466.
- Merritt SM, Ilgen DR. Not all trust is created equal: dispositional and history-based trust in human–automation interactions. *Human Factors*. 2008;50(2):194–210.
- Merritt SM, Heimbaugh H, LaChapell J, Lee D. I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*. 2013;55(3):520–534. doi.org/10.1177/0018720812465081.
- Meyer J. Effects of warning validity and proximity on responses to warnings. *Human Factors*. 2001;43(4):563–572. doi:10.1518/001872001775870395.

- Meyer J. Conceptual issues in the study of dynamic hazard warnings. *Human Factors*. 2004;46(2):196–204. doi:10.1518/hfes.46.2.196.37335.
- Meyer J, Wiczorek R, Günzler T. Measures of reliance and compliance in aided visual scanning. *Human Factors*. 2014;56:840–849. doi:10.1177/0018720813512865.
- Mikus CR, Earnest CP, Blair SN, Church TS. Heart rate and exercise intensity during training: observations from the DREW Study. *British Journal of Sports Medicine*. 2009;43(10):750–755.
- Milner A, Han Seong D, Brewer R, Baker AL, Krausman A, Chhan D, Thomson RH, Rovira E, Schaefer KE. Identifying new team trust and team cohesion metrics that support future human–autonomy teams. In: Cassenti D, Scataglini S, Rajulu S, Wright J, editors. *Advances in Simulation and Digital Human Modelling*. AHFE 2020. Springer, Cham 2020. (Advances in Intelligent Systems and Computing; vol. 1206). doi.org/10.1007/978-3-030-51064-0_12.
- Mitkidis P, McGraw JJ, Roepstorff A, Wallot S. Building trust: heart rate synchrony and arousal during joint action increased by public goods game. *Physiology and Behavior*. 2015;149:101–106.
- Mønster D, Hakonsson DD, Eskildsen JK, Wallot S. Physiological evidence of interpersonal dynamics in a cooperative production task. *Physiology and Behavior*. 2016;156:24–34.
- Montague E, Xu J, Chiou E. Shared experiences of technology and trust: an experimental study of physiological compliance between active and passive users in technology-mediated collaborative encounters. *IEEE Transactions on Human–Machine Systems*. 2014;44(5):614–624.
- Mooradian T, Renzl B, Matzler K. Who trusts? Personality, trust and knowledge sharing. *Management Learning*. 2006;37(4):523–540.
- Morrow PB, Fiore SM. Supporting human–robot teams in social dynamicism: an overview of the metaphoric inference framework. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2012 Oct 22–26; Boston, MA. SAGE Publications; c2012;56(1):1718–1722.
- Mower EP, Feil-Seifer D, Mataric M, Narayanan S. Investigating implicit cues for user state estimation in human–robot interaction using physiological measurements. *IEEE International Conference on Robot and Human Interactive Communication*; 2007. p. 1125–1130.

- Muir BM. Trust between humans and machines, and the design of decision aids. *International Journal of Man–Machine Studies*. 1987;27(5–6):527–539.
- Muir BM, Moray N. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*. 1996;39(3):429–460. doi:10.1080/00140139608964474.
- Neubauer C, Gremillion G, Perelman B, La Fleur C, Metcalfe J, Schaefer-Lay K. How analysis of facial expressions explain affective state and trust-based decisions during interaction with autonomy aids. *Proceedings of the 3rd International Conference on Intelligent Human Systems Integration: Integrating People and Intelligent Systems*; 2020 Feb 19–20; Modena, Italy.
- Newton D. The perception and coupling of behavior waves. In: Vallacher RR, Nowak A, editors. *Dynamical Systems in Social Psychology*. Academic Press; 1994. p. 139–167.
- Nguyen TLT, Kiang TB. Understanding shared space for informal interaction among geographically distributed teams. In: *Proceedings of CAAD Futures*; 2011. p. 41–54.
- Nicholson S. The relevant and reliable language theory: developing a language measure of trust for online groups [doctoral dissertation]. Lancaster University; 2017.
- Niederhoffer KG, Pennebaker JW. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*. 2002;21(4): 337–360.
- Norscia I, Palagi E. Yawn contagion and empathy in *Homo sapiens*. *PLoS ONE*. 2011;6(12):Article e28472.
- O’hara K, Kjeldskov J, Paay J. Blended interaction spaces for distributed team collaboration. *ACM Transactions on Computer–Human Interaction (TOCHI)*; 2011;18(1):1–28.
- Oleson KE, Hancock PA, Billings DR, Schesser CD. Trust in unmanned aerial systems: a synthetic, distributed trust model. *16th International Symposium on Aviation Psychology*; 2011. p. 469–474.
- Ososky S. Influence of task-role mental models on human interpretation of robot motion behavior [doctoral dissertation]. University of Central Florida; 2013.
- Ososky S, Schuster D, Phillips E, Jentsch F. Building appropriate trust in human–robot teams. *AAAI Spring Symposium: Trust and Autonomous Systems*; 2013 Mar 25–27; Palo Alto, CA. p .60–65.

- Pak R, Crumley-Branyon JJ, de Visser EJ, Rovira E. Factors that affect younger and older adults' causal attributions of robot behaviour. *Ergonomics*. 2020;63(4):421–439.
- Parasuraman R, Molloy R, Singh IL. Performance consequences of automation-induced complacency. *Int J Av Psych*. 1993; 3(1):1–23.
- Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. *Human Factors*. 1997;39(2):230–253.
- Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. *IEEE Trans Sys Man Cyber Part A: Sys Hum*. 2000; 30(3):286–297.
- Parasuraman R, Cosenzo KA, de Visser E. Adaptive automation for human supervision of multiple uninhabited vehicles: effects on change detection, situation awareness, and mental workload. *Military Psychology*. 2009;21(2):270.
- Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. *Human Factors*. 2010; 52(3):381–410.
- Pelikan HR, Cheatle A, Jung MF, Jackson SJ. Operating at a distance—how a teleoperated surgical robot reconfigures teamwork in the operating room. *Proceedings of the ACM on Human–Computer Interaction*. 2018;2(CSCW):1–28.
- Pennebaker JW, Booth RJ, Francis ME. Linguistic inquiry and word count: LIWC. *liwc.net*; 2007. 135.
- Perelman BS, Evans AW, Schaefer KE. Where do you think you're going? Characterizing spatial mental models from planned routes. *J Hum Robot Interact*. 2020 Sep;9(4):Article 23. 55 pages. doi.org/10.1145/3385008.
- Phillips E, Ososky S, Grove J, Jentsch F. From tools to teammates: toward the development of appropriate mental models for intelligent robots. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2011 Sep 26–30; Las Vegas, NV. SAGE Publications; c2011;55(1):1491–1495. doi.org/10.1177%2F1071181311551310.
- Pilisuk M, Skolnick P. Inducing trust: a test of the Osgood proposal. *Journal of Personality and Social Psychology*. 1968;8(2p1):121.
- Pilisuk M, Kiritz S, Clampitt S. Undoing deadlocks of distrust: hip Berkeley students and the ROTC. *Journal of Conflict Resolution*. 1971;15(1):81–95.

- Poole A, Ball L. Eye tracking in human–computer interaction and usability research: Current status and future prospects. 2006.
- Reed RG, Randall, AK, Post JH, Butler EA. Partner influence and in-phase versus anti-phase physiological linkage in romantic couples. *International Journal of Psychophysiology*; 2013;88:309–316.
- Richardson DC, Dale R. Looking to understand: the coupling between speakers’ and listeners’ eye movements and its relationship to discourse comprehension. *Cognitive Science*. 2005;29:1045–1060.
- Rieh SY, Belkin N. Understanding judgment of information quality and cognitive authority in the WWW. *Journal of The American Society for Information Science and Technology - JASIS*. 1998;35.
- Robinette P, Li W, Allen R, Howard A, Wagner AR. Overtrust of robots in emergency evacuation scenarios. *IEEE International Conference on Human–Robot Interaction*; 2016. doi:10.1109/HRI.2016.7451740.
- Ross JM, Szalma, JL, Hancock PA. Empirical examination of trust in automation across multiple agents in a search and rescue operation. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; 2007. SAGE Publications. Vol. 51; No. 22; p. 1501–1505.
- Rotter JB. A new scale for the measurement of interpersonal trust. *Journal of Personality*. 1967;35:651–665.
- Rotter JB. Generalized expectancies of interpersonal trust. *American Psychologist*. 1971;26(5):443–452.
- Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: a cross-discipline view of trust. *Academy of Management Review*. 1998;23. 10.5465/AMR.1998.926617.
- Russell S, Norvig P. *Artificial intelligence: a modern approach*. 3rd Ed. Prentice Hall; 2009.
- Salas E, Dickinson TL, Converse SA, Tannenbaum SI. Toward an understanding of team performance and training. In: Swezey RW, Salas E, editors. *Teams: their training and performance*. Ablex Publishing; 1992. p. 3–29.
- Sarker S, Ahuja M, Sarker S, Kirkeby S. The role of communication and trust in global virtual teams: a social network perspective. *Journal of Management Information Systems*. 2011;28(1);273–310.

- Saunderson S, Nejat G. How robots influence humans: a survey of nonverbal communication in social human–robot interaction. *International Journal of Social Robotics*. 2019;11(4):575–608.
- Savolainen T, Ikonen M. Process dynamics of trust development: exploring emergence in team context. *Trust, Organizations, and Social Organizations: Studying Trust as Process Within and Between Organizations*. Edward Elgar Publishing; 2016. p. 231–256.
- Schaefer KE. The perception and measurement of human–robot trust [dissertation]. University of Central Florida; 2013.
- Schaefer KE, Sanders TL, Yordon RE, Billings DR, Hancock PA. Classification of robot form: factors predicting perceived trustworthiness. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*; 2012; Santa Monica, CA. Human Factors and Ergonomics Society. p. 1548–1552.
- Schaefer KE, Evans AW, Hill SG. Command and control in network-centric operations: trust and robot autonomy. *Proceedings of the 20th International Command and Control Research and Technology Symposium*; 2015; Annapolis, MD. Command and Control Research Program. <http://www.dodccrp-test.org/2015>.
- Schaefer KE, Chen JYC, Szalma JL, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. *Hum Fact*. 2016;58:377–400.
- Schaefer KE, Straub ER, Chen JYC, Putney J, Evans AW. Communicating intent to develop shared situation awareness and engender trust in human–agent teams. *Cognitive Systems Research*. 2017;46:26–39.
- Schaefer KE, Aksaray D, Wright JL, Roy N. Challenges with addressing the issue of context within AI and human–robot teaming. In: Lawless W, Mittu R, Sofge D, editors. *Computational context: the value, theory and application of context with AI*. CRC Press; 2018a. doi.org/10.1201/9780429453151.
- Schaefer KE, Fouts AN, Straub ER. Applied robotics for installations and base operations: user perceptions of a driverless vehicle at Fort Bragg. *Army Research Laboratory (US)*; 2018b. Report No.: ARL-TR- 8265.
- Schaefer KE, Oh J, Aksaray D, Barber D. Integrating context into artificial intelligence: research from the robotics collaborative technology alliance. *AI Magazine*. 2019a;40(3):28–40. doi.org/10.1609/aimag.v40i3.2865.

- Schaefer KE, Baker AL, Brewer RW, Patton DJ, Canady JD, Metcalfe JS. Assessing multi-agent human–autonomy teams: US Army Robotic Wingman gunnery operations. *Proc. SPIE 10982, Micro- and Nanotechnology Sensors, Systems, and Applications XI Conference*; 2019 Apr 14–18; Baltimore, MD. 2019b;109822B. doi: 10.1117/12.2519302.
- Schaefer KE, Perelman BS, Gremillion GM, Marathe AR, Metcalfe JS. A roadmap for developing team trust for human–autonomy teams. In: Lyons J, Nam C, editors. *Trust in Human–Robot Interaction: Research and Applications*. Elsevier; 2021.
- Schmidt RC, O’Brien B. Modeling interpersonal coordination dynamics: implications for a dynamical theory of developing systems. In: Molenaar PC, Newell K, editors. *Dynamics Systems and Development: Beyond the Metaphor*. Erlbaum; 1998. p. 221–240.
- Schneider TR, Jessup SA, Stokes C, Rivers S, Lohani M, McCoy M. The influence of trust propensity on behavioral trust. Poster session presented at the meeting of Association for Psychological Society; 2017 May; Boston, MA.
- Schoorman FD, Mayer RC, Davis JH. An integrative model of organizational trust: past, present, and future. *The Academy of Management Review*. 2007;32(2):344–354. doi.org/10.2307/20159304.
- Schweitzer ME, Hershey JC, Bradlow ET. Promises and lies: restoring violated trust. *Organizational Behavior and Human Decision Processes*. 2006;101:1–19.
- Scissors LE, Gill AJ, Geraghty K, Gergle D. In CMC we trust: the role of similarity. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*; 2009 Apr. p. 527–536.
- Sebanz N, Shiffrar M. Bodily bonds: effects of social context on ideomotor movements. In: Haggard P, Rossetti Y, Kawato M, editors. *Sensorimotor Foundations of Higher Cognition: Attention and Performance*; vol. 22:267–292. Oxford University Press; 2007.
- Serva MA, Fuller MA, Mayer RC. The reciprocal nature of trust: a longitudinal study of interacting teams. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*. 2005;26(6):625–648.
- Shi Y, Ruiz N, Taib R, Choi E, Chen F. Galvanic skin response (GSR) as an index of cognitive load. CHI EA ‘07 Extended Abstracts on Human Factors in Computing Systems. 2007;2651–2656.

- Shockley K, Santana MV, Fowler CA. Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*. 2003;29:326–332.
- Simons TL, Peterson RS. Task conflict and relationship conflict in top management teams: the pivotal role of intragroup trust [electronic version]. Cornell University, School of Hospitality Administration; 2000. <http://scholarship.sha.cornell.edu/articles/719>.
- Singh IL, Molloy R, Parasuraman R. Automation-induced “complacency”: development of the complacency-potential rating scale. *Int J Aviat Psychol*. 1993;3:111–122.
- Spain RD, Bliss JP. The effect of sonification display pulse rate and reliability on operator trust and perceived workload during a simulated patient monitoring task. *Ergonomics*. 2008;51(9):1320–1337. doi:10.1080/00140130802120234.
- Sperber D, Clément F, Heintz C, Mascaro O, Mercier H, Origgi G, Wilson D. Epistemic vigilance. *Mind and Language*. 2010;25(4):359–393.
- Sperrle F, Schlegel U, El-Assady M, Keim D. Human trust modeling for bias mitigation in artificial intelligence. In: *ACM CHI 2019 Workshop: Where is the Human? Bridging the Gap Between AI and HCI*; 2019.
- Staples DS, Webster J. Exploring the effects of trust, task interdependence and virtualness on knowledge sharing in teams. *Information Systems Journal*. 2008;18(6):617–640. doi:10.1111/j.1365-2575.2007.00244.x.
- Steain A, Stanton CJ, Stevens CJ. The black sheep effect: the case of the deviant ingroup robot. *PloS one*. 2019;14(10):e0222975.
- Stephens JP, Heaphy ED, Carmeli A, Spreitzer GM, Dutton JE. Relationship quality and virtuousness: emotional carrying capacity as a source of individual and team resilience. *Journal of Applied Behavioral Science*. 2013;49:13–41.
- Strang AJ, Funke GJ, Russell SM, Dukes AW, Middendorf MS. Physio-behavioral coupling in a cooperative team task: contributors and relations. *Journal of Experimental Psychology: Human Perception and Performance*; 2013;40:145–158.
- Stuck R. Understanding dimensions of trust between older adults and human or robot care providers [master’s thesis]. Georgia Institute of Technology; 2017.
- Szalma JL, Taylor GS. Individual differences in response to automation: the five factor model of personality. *Journal of Experimental Psychology*. 2011;Applied 17(2):71.

- Thompson LF, Gillan DJ. Social factors in human-robot interaction. In: Barnes M, Jentsch F, editors. *Human–Robot Interactions in Future Military Operations*. Ashgate; 2010. p. 67–82.
- Thurman SM, Wasylyshyn N, Roy H, Lieberman G, Garcia JO, Asturias A, Okafor G, Elliott JC, Giesbrecht B, Grafton ST, et al. Individual differences in compliance and agreement for sleep logs and wrist actigraphy: a longitudinal study of naturalistic sleep in healthy adults. *PloS ONE*. 2018;13(1):e0191883.
- Tiferes J, Hussein AA, Bisantz A, Kozlowski JD, Sharif MA, Winder NM, Ahmad N, Allers J, Cavuoto LA, Guru KA. The loud surgeon behind the console: understanding team activities during robot-assisted surgery. *Journal of Surgical Education*. 2016;73(3):504–512. doi:10.1016/j.jsurg.2015.12.009.
- Tognoli E, Lagarde J, DeGuzman GC, Kelso JAS. The phi complex as a neuromarker of human social coordination. *PNAS, Proceedings of the National Academy of Sciences of the United States of America*. 2007;104:8190–8195.
- Tomlinson EC, Dineen BR, Lewicki RJ. The road to reconciliation: antecedents of victim willingness to reconcile following a broken promise. *J Manage*. 2004 Apr;30(2):165–187. doi.org/10.1016/j.jm.2003.01.003.
- Tuckman BW, Jensen MA. Stages of small-group development revisited. *Group and Organization Studies*. 1977;2(4):419–427. doi.org/10.1177/105960117700200404.
- van Ulzen NR, Lamothe CJC, Daffertshofer A, Semin GR, Beek PJ. Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience Letters*. 2008;432:88–93.
- Wachbroit R. Reliance and reliability: the problem of information on the internet. Report from the Institute for Philosophy and Public Policy. 2000;20(4):9–15.
- Wang L, Jamieson GA, Hollands JG. Trust and reliance on an automated combat identification system. *Human Factors*. 2009;51(3):281–291.
- Wiczorek R, Manzey D. Is operators' compliance with alarm systems a product of rational consideration? *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*; 2010; San Francisco, CA. p. 1722–1726.
- Wilkinson GL, Bennett LT, Oliver KM. Evaluation criteria and indicators of quality for internet resources. *Educational Technology*. 1997;37(3):52–59.
- Wright J, Sanders T, Hancock PA. Identifying the role of attributions in human perceptions of robots. *Proceedings of the Human Factors and Ergonomics*

- Society Annual Meeting; 2013 Sep. SAGE Publications. Vol. 57; No. 1; p. 1288–1292.
- Wright JL, Chen JYC, Barnes MJ, Hancock PA. Agent reasoning transparency: the influence of information level on automation-induced complacency. Army Research Laboratory (US); 2017. Report No.: ARL-TR-8044.
- Wright JL, Chen JYC, Barnes MJ. Human–automation interaction for multiple robot control: the effect of varying automation assistance and individual differences on operator performance. *Ergonomics*. 2018;61(8):1033–1045.
- Wright JL, Chen JY, Lakhmani SG. Agent transparency and reliability in human–robot interaction: the influence on user confidence and perceived reliability. *IEEE Transactions on Human–Machine Systems*. 2020;50(3):254–263. doi:10.1109/THMS.2019.2925717.
- Yagoda R, Gillan D. WHAT! You want me to trust a ROBOT? The development of a human robot interaction (HRI) trust scale. *International Journal of Social Robotics*. 2012;4:235–248.
- Yamagishi T, Yamagishi M. Trust and commitment in the United States and Japan. *Motivation and Emotion*. 1994;18(2):129–166.
- Yamagishi T, Cook KS, Watabe M. Uncertainty, trust, and commitment formation in the United States and Japan. *American Journal of Sociology*. 1998;104(1):165–194.
- Yamagishi T, Kikuchi M, Kosugi M. Trust, gullibility, and social intelligence. *Asian Journal of Social Psychology*. 1999;2(1):145–161.
- Yamamoto Y, Tezuka T, Jatowt A, Tanaka K. Supporting judgment of fact trustworthiness considering temporal and sentimental aspects. *International Conference on Web Information Systems Engineering*. Springer, Berlin, Heidelberg; 2008 Sep. p. 206–220.
- Yuki M, Maddux WW, Brewer MB, Takemura K. Cross-cultural differences in relationship- and group-based trust. *Personality and Social Psychology Bulletin*. 2005;31(1):48–62.
- Zaheer A, Harris J. Interorganizational trust. In: Shenkar O, Reuer J; editors. *Handbook of Strategic Alliances*. 2006;169–197.
- Zhang S, Zhong H. Mining users trust from e-commerce reviews based on sentiment similarity analysis. *IEEE Access*. 2019;7:13523–13535.

- Zhou J, Sun J, Chen F, Wang Y, Taib R, Khawli A, Li Z. Measurable decision making with GSR and pupillary analysis for intelligent user interfaces. *ACM Transactions on Computer–Human Interaction*. 2015;21(6):1–23.
- Zhou J, Hu H, Li Z, Yu K, Chen F. Physiological indicators for user trust in machine learning with influence enhanced fact-checking. In: *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, Cham; 2019 Aug. p. 94–113.

List of Symbols, Abbreviations, and Acronyms

AI	artificial intelligence
a.k.a.	also known as
ARL	Army Research Laboratory
AU	action unit
BVP	blood volume pulse
CPRS	Complacency Potential Rating Scale
DEVCOM	US Army Combat Capabilities Development Command
EDA	electrodermal activity
EEG	electroencephalography
FACS	Facial Action Coding System
fEMG	facial electromyographic activity
HCI	human–computer interaction
HR	heart rate
HRI	Human–Robot Interaction
HRV	heart rate variability
IA	intelligent agent
IBI	interbeat interval
IMT	Integrated Model of Trust
ITS	Interpersonal Trust Scale
LDA	Latent Dirichlet Allocation
LIWC	Linguistic Inquiry and Word Count
LRAS	Long-Range Advance Scout Surveillance System Operator
LSA	Latent Semantic Analysis
LSM	Language Style Matching
N/A	not applicable
NASA	National Aeronautics and Space Administration

PBC	physiobehavioral coupling
PTS	Propensity to Trust Survey
PTT	Propensity to Trust in Technology
RQA	recurrence quantification analysis
RVG	Robotic Vehicle Gunner
TLX	Task Load Index
VC	Vehicle Commander

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 DEVCOM ARL
(PDF) FCDD RLD DCI
TECH LIB

1 DEVCOM ARL
(PDF) FCDD RLH B
T DAVIS
BLDG 5400 RM C242
REDSTONE ARSENAL AL
35898-7290

1 DEVCOM ARL
(PDF) FCDD HSI
J THOMAS
6662 GUNNER CIRCLE
ABERDEEN PROVING
GROUND MD
21005-5201

1 USN ONR
(PDF) ONR CODE 341 J TANGNEY
875 N RANDOLPH STREET
BLDG 87
ARLINGTON VA 22203-1986

1 USA NSRDEC
(PDF) RDNS D D TAMILIO
10 GENERAL GREENE AVE
NATICK MA 01760-2642

1 OSD OUSD ATL
(PDF) HPT&B B PETRO
4800 MARK CENTER DRIVE
SUITE 17E08
ALEXANDRIA VA 22350

ABERDEEN PROVING GROUND

23 DEVCOM ARL
(PDF) FCDD RLH
J LANE
Y-S CHEN
P FRANASZCZUK
K MCDOWELL
FCDD RLH F
J GASTON
K OIE
FCDD RLH FA
AW EVANS
G BOYKIN
C NEUBAUER
D FORSTER
SM FITZHUGH
JS METCALFE
FCDD RLH FB
J GARCIA (A)
H ROY
FCDD RLH FC
J TOURYAN (A)
T ROHALY
FCDD RLH FD
A MARATHE
A KRAUSMAN
S LAKHMANI
AL BAKER
G GREMILLION
JL WRIGHT
FCDD RLC IS
KE SCHAEFER