



# Approaches for assessing communication in human-autonomy teams

Anthony L. Baker<sup>1</sup> · Sean M. Fitzhugh<sup>1</sup> · Lixiao Huang<sup>2</sup> · Daniel E. Forster<sup>1</sup> · Angelique Scharine<sup>1</sup> · Catherine Neubauer<sup>1</sup> · Glenn Lematta<sup>2</sup> · Shawaiz Bhatti<sup>2</sup> · Craig J. Johnson<sup>2</sup> · Andrea Krausman<sup>1</sup> · Eric Holder<sup>1</sup> · Kristin E. Schaefer<sup>1</sup> · Nancy J. Cooke<sup>2</sup>

Received: 16 July 2020 / Accepted: 26 January 2021  
© The Author(s) 2021

## Abstract

Evaluation of team communication can provide critical insights into team dynamics, cohesion, trust, and performance on joint tasks. Although many communication-based measures have been tested and validated for human teams, this review article extends this research by identifying key approaches specific to human-autonomy teams. It is not possible to identify all approaches for all situations, though the following seem to generalize and support multi-size teams and a variety of military operations. Therefore, this article will outline several key approaches to assessing communication, associated data requirements, example applications, verification of methods through HAT use cases, and lessons learned, where applicable. Some approaches are based on the structure of team communication; others draw from dynamical systems theory to consider perspectives across different timescales; other approaches leverage features of team members' voices or facial expressions to detect emotional states that can provide windows into other workings of the team; still others consider the content of communication to produce insights. Taken together, these approaches comprise a varied toolkit for deriving critical information about how team interactions affect, and are affected by, coordination, trust, cohesion, and performance outcomes. Future research directions describe four critical areas for further study of communication in human-autonomy teams.

**Keywords** Communication assessment · Metrics and measurement · Human-autonomy teams · Team performance · Trust · Cohesion

## 1 Introduction

Communication is the primary mechanism through which teams exchange information, coordinate actions, develop strategies, and execute plans (Salas et al. 2005). However, there is major gap in our understanding regarding best practices for how to effectively study and assess communication in human-autonomy teams. It is a critical need because communication, regardless of its modality, forms the foundation of effective teamwork and is a key factor in understanding the success or failure of teams (Mesmer-Magnus and DeChurch 2009; Salas et al. 2008). Thus, evaluation of communication

can provide critical insights into elements of team cohesion, team trust, and even help describe the reasoning behind team performance. Although it is clear that communication is fundamental to human-autonomy team success, it is currently unclear what communication analyses are most useful for understanding team states and outcomes in various scenarios, tasks, and team configurations. Therefore, the purpose of this article is threefold: (1) identify key measures of communication for human-autonomy teams; (2) describe how the measures can be applied; (3) discuss how those measures are associated with cohesion, trust, performance, or other outcomes in human-autonomy teams. The following subsections provide an introduction to the space of human-autonomy teaming, team trust and cohesion, and a background on communication research.

### 1.1 Defining human-autonomy teams

Before delving into communication assessments, it is important to understand the terminology associated with human-

✉ Anthony L. Baker  
TLBaker27@gmail.com

<sup>1</sup> DEVCOM Army Research Laboratory, Aberdeen Proving Ground, MD, USA

<sup>2</sup> Center for Human, Artificial Intelligence and Robot Teaming, Arizona State University, Mesa, AZ, USA

autonomy teams. For this scope, human-autonomy teams involve one or more humans and one or more autonomy-enabled systems, or *intelligent agents* (IA), who coordinate and work interdependently over time in order to reach a common goal or complete a task (McNeese et al. 2018). An IA is an entity that possesses the capability to sense, observe and act on the environment, and intelligently respond to unexpected, dynamic events (Johnson et al. 2018; Mercado et al. 2016). In future military operations, there is the potential for interaction with a wide variety of embodied IAs—from small ground, air, surface or subsurface unmanned vehicles or robots to large weaponized unmanned vehicles, troop or cargo carriers, and even computer-based IAs that assist with decision-making, planning, or communication. The reason these are called *autonomy-enabled systems* is to denote the multiple types of autonomy (e.g., mobility, planning, decision-making, weapon system, and intelligent computer agent) that often make up an embodied IA, such as a robot or unmanned vehicle. Each type of autonomy could have a variable *level of automation* which refers to how functions are distributed among humans and the system or agent (Parasuraman et al. 2000; Lewis et al. 2018). Seminal work by Sheridan and Verplank (1978) suggests a continuum of automation levels ranging from where the user performs the task manually with no assistance from the agent, to full automation in which the system or agent governs its own actions, goals, and decisions. Systems with fully autonomous capabilities are, in a sense, more adaptable, capable, independent, and interdependent than traditional automation and can arguably represent evolved forms of automation (Endsley 2015; Endsley 2017; Hancock 2017). Although IAs will continue to advance and acquire more capabilities, the goal of using these systems is not to replace the human, or to provide more “complex tools” for humans to use, but rather to effectively “team” autonomy with humans so they achieve synergy and are able to work cooperatively in complex, dynamic environments (Phillips et al. 2011; Johnson and Vera 2019). For example, the US military is researching how human-autonomy teams can maintain overmatch in unpredictable adversarial environments including intelligence, surveillance, and reconnaissance (Harris and Barber 2014; Kattoju et al. 2016) as well as lethality (Brewer et al. 2018; Schaefer et al. 2019a) while reducing the cognitive burden on Soldiers. A significant challenge is integrating humans and autonomous systems into interdependent, heterogeneous teams that are able to achieve appropriate levels of team trust and cohesion.

## 1.2 Team trust and cohesion

As we move toward integration of autonomous IAs working collaboratively with human team members, it becomes imperative to calibrate the trust in the team and formulate an appropriate level of team cohesion for effective team performance to ensue. With team trust, team members place less emphasis

on personal interests, minimizing the effects of uncertainty and vulnerability toward their teammates (De Jong and Elfring 2010). This allows team members to take risks that facilitate cooperation and overall effectiveness (Colquitt et al. 2007). A lack of trust, in contrast, causes breakdowns in goal-directed teamwork and causes members to place greater emphasis on personal interests over team interests (Joshi et al. 2008). When integrating autonomy into the team, it is important to understand that human-autonomy trust functions differently from trust between humans (Baker et al. 2018), but well-calibrated trust in autonomy is still positively correlated with team performance (McNeese et al. 2019). To understand and contextualize trust and performance in a human-autonomy team, then, it is critical to evaluate the communication of the team, as this provides a window into the team’s coordination, information sharing, decision-making, and more (Baker et al. 2019; Baker et al. 2020a, b; Schaefer et al. 2019a).

Team cohesion, while described in many different ways, generally comprises attraction or bonding within a group that emerges as a function of the group’s shared experiences, identities, or goals (Salas et al. 2015). Research suggests that how team members feel about a task influences teamwork (Salas and Cannon-Bowers 2000), and that team cohesion, or the ability to work as a team, has a direct impact on team behaviors and performance (Beal et al. 2003). In military contexts, research has demonstrated a positive relationship between cohesion and team performance (Ahronson and Cameron 2007; Oliver et al. 1999) that is mirrored in across academic, organizational, and other contexts as well (Beal et al. 2003; Chiochio and Essiembre 2009). Because team cohesion and trust are emergent states that relate to performance, it will be key to understand how they emerge and evolve in human-autonomy teams, especially as a function of the team’s communication.

The development of more effective human-autonomy teams will require a focus on the interaction patterns among team members. To perform optimally, these human-autonomy teams will need to exhibit a shared understanding of mission-relevant aspects of their goals, tasks, intentions, teamwork, and environment (Ososky et al. 2012) that can develop through the kinds of dynamic, adaptive interactions that are characteristic of human teams (Sycara and Sukthankar 2006). Thus, analysis of how information is communicated and what is communicated can provide critical insights into team dynamics needed to calibrate trust. For example, Dzindolet et al. (2003) found that when people are provided explanations for errors made by an automated decision aid, they show a closer alignment between the system’s reliability and their trust in the system; in contrast, when people do not receive such explanations, there is poorer trust calibration, sometimes resulting in distrust of a reliable aid. In addition, when human team members are informed about the reliability limitations of the IA, they are able to calibrate appropriate use of the system

at the appropriate time (e.g., automated combat identification system; Wang et al. 2009). These results indicate that a static understanding of an autonomy's reliability is not a viable substitute for timely and relevant communication, which is critical to increasing human-autonomy trust and performance (Schaefer and Straub 2016).

### 1.3 Communication

Communication involves the transmission of information both within and across teams, and as such there are a multitude of factors that affect how well a team communicates. Within human-autonomy teams, the autonomy introduces unique capabilities and challenges into a team. Much of that challenge stems from how interactions with the autonomy affect communication patterns. People use communication as a way to share information (Mesmer-Magnus et al. 2011), set goals (Marks et al. 2001), self-correct (Salas et al. 2008), and engage in teamwork. However, many autonomous systems in task-oriented team environments cannot yet engage in naturalistic communication, although research is working toward this end (Bisk et al. 2016; Marge et al. 2016; Demir et al. 2020; Thomason et al. 2020). Improving the capability of autonomy to more effectively communicate intentions to human team members, such as designing user displays to allow IAs to communicate intent, are key challenges in human-autonomy team communication (Schaefer et al. 2017). Communication in human-autonomy teams may also rely on alternative methods other than speech to communicate. For example, research has shown the benefits of using visual, gesture, tactile or haptic feedback, or other non-verbal auditory cues to initiate bidirectional communication (Barber et al. 2015; Hill 2017; Elliott et al. 2016).

Communication theories and models provide critical foundations for understanding human-autonomy team communication (see Baker et al. 2019). The information-theory model of Shannon (1948) laid important groundwork, conceptualizing the transfer of information from a sender to a receiver. Berlo (1960) built on this by highlighting the importance of the communication channel. Sacks et al. (1974) pioneered a system for breaking down a conversation into sequences of turn-taking. Clark and Brennan (1991) outlined how we use language to establish a common understanding, in a landmark manuscript that tied communication to cognition. Seminal works such as these reveal historical foundations that can contribute to a clearer understanding of communication in human-autonomy teams, especially as communication assessment approaches are developed to address key research gaps.

## 2 Communication assessment approaches for human-autonomy teams

As there are many approaches for communication assessment, it is important to weigh the characteristics of each assessment in order to elicit the most useful data from a given human-autonomy team interaction. This section describes eleven approaches for assessing communication in human-autonomy teams. The first four approaches are based on the structure of team interactions. The following two approaches draw from dynamical systems theory to consider perspectives across different timescales. After that, two approaches leverage features of team members' voices or facial expressions to detect emotional states that can provide windows into other workings of the team. The final three approaches fall under the umbrella of linguistic synchrony, offering insights into the content of a team's communication. Although this is not an exhaustive list of approaches that can be used for assessing human-autonomy team communication, it provides many ways to understand team communication using different types of input data.

Discussion of each communication assessment approach comprises the following four areas: (1) a description of the approach, (2) the method by which the approach is carried out, (3) applications or considerations for using the approach, and (4) a description of current directions in developing and applying these approaches to human-autonomy teams. Table 1 provides a brief overview of each of the approaches described in this paper and describes the types of data or team interactions that are best studied with each approach. Following Table 1, we begin discussing communication assessment approaches by describing the four approaches based on the structure of team communication. These approaches primarily rely on understanding how information is sent and received.

### 2.1 Structural analysis

#### 2.1.1 Aggregate communication flow modeling

Team communication patterns provide a window into how a team completes tasks, achieves goals, and coordinates information (Sacks et al. 1974; Kiekel et al. 2001; Tiferes et al. 2016). Communication flow, or the measurement of how information is passed throughout a team, can allow one to evaluate a team's coordination behaviors (Fischer et al. 2007). When aggregated, communication flow can be visually represented to reveal how teams share information, which in turn may relate to how teams experience trust and team cohesion. These insights will be important to the development of human-autonomy teams as they will aid us in understanding how communication affects, and is affected by, intelligent systems, team structures, and task demands. For example, Fischer et al. (2007) demonstrated structural differences in

**Table 1** Approaches for assessing communication in human-autonomy teams

Approach: Structural analysis	Description	Useful for assessing...	Output	Citations
Aggregate communication flow	Proportion of communication accounted for by each sender-receiver pair	Communication interactions with a clear sender and a receiver (e.g., speech, chat messages, etc.)	Communication flow diagrams	Fischer et al. (2007); Tiferes et al. (2016); Schaefer et al. 2019a
Social network analysis	Structural connections among a set of actors	Interactions with a clear sender and receiver (e.g., speech, chat messages, or any directed interaction).	Descriptive measures of connectedness	Wasserman and Faust (1994); Katz et al. (2004); Boghrati et al. (2018)
Relational event models	Model-based approach for analyzing dynamic networks	Interactions with a clear sender and receiver (e.g., speech, chat messages, or any directed interaction).	Model coefficients	Butts (2008); Leenders et al. (2016)
Anticipatory information pushing (AIP)	Team member anticipates informational needs and preemptively pushes information	Interactions with a clear sender and receiver (e.g., speech, chat messages, or any directed interaction).	Frequency of anticipatory information pushing behaviors	Demir et al. (2019); Demir et al. 2017
Approach: Dynamical systems				
Distributed dynamic approach of team cognition	Customizable, holistic approach for assessing team dynamics over time	Team cognition among multiple members and varying stakeholders over time	Team communication network patterns across phases	Cooke (2015); Huang et al. (2020a)
Exploratory communication analysis	Approach for identifying novel communication within dynamic patterns	Interactions with a clear sender and receiver (e.g., speech, chat messages, or any directed interaction).	Variability in exploratory communication behaviors of teams	Gorman et al. (2019); Amazeen (2018); Demir et al. (2018)
Approach: Emotional states				
Facial expression analysis	Detecting emotional states from facial features	Facial expressions and emotional affect	Evidence of action units and corresponding facial expressions	Neubauer et al. (2017); Ekman and Friesen (1978)
Vocal feature assessment using neural networks	Identifying emotions based on vocal features	Voice, speech, and emotional affect	Emotional state classification	Casale et al. 2008; Koolagudi and Rao (2012)
Approach: Linguistic synchrony				
Language style matching	Proportion of function words used in a conversation	Dialogue (speech or text) between two or more speakers	Similarity in rate of word category usage	Niederhoffer and Pennebaker (2002); Gonzales et al. (2010)
Latent semantic similarity	Constructs word meanings based on co-occurrences with other words	Dialogue (speech or text) between two or more speakers	Coherence between speakers' utterances in semantic space	Landauer and Dumais (1997); Foltz et al. (2003)
Conversation level syntax similarity metric (CASSIM)	Analyzes syntactic structures by automatically parsing sentences	Dialogue (speech or text) between two or more speakers	Estimate of syntactic similarity	Boghrati et al. (2018); Boghrati and Dehghani (2018)

the communication patterns of successful and unsuccessful teams, revealing that successful teams in search-and-rescue operations had a more equal distribution of communication, whereas unsuccessful teams tended to involve a few team members dominating the discourse. This finding was echoed by Hung and Gatica-Perez (2010), who found that highly cohesive small groups demonstrated a balanced amount of discussion among team members, whereas teams with one person dominating the conversation demonstrated little cohesion. This suggests sender-receiver information can be used to predict the cohesiveness of a team. Research will need to investigate the extent to which these communication links can relate to perceptions of team cohesion in both novel and intact human-autonomy teams. Furthermore, it will be important to understand how team communication patterns relate to their performance in a given scenario or context. This will shed additional light on how their communication can affect cohesion, as well as trust, given that some aspects of both cohesion and trust are affected by team outcomes.

There are two related advantages to evaluating team communication flow in aggregate. First, depending on the communication modality, it can be objectively derived. For example, in the case of event logs that contain a sender, receiver, and timestamp, this data (and accompanying aggregate flow maps) can be instantly collected and generated for review. Second, it is easily interpretable. Aggregating a team's communication flow over time can reveal a clear picture of the proportion of team interactions accounted for by each sender-receiver pair, providing a window into the relative importance of each pair.

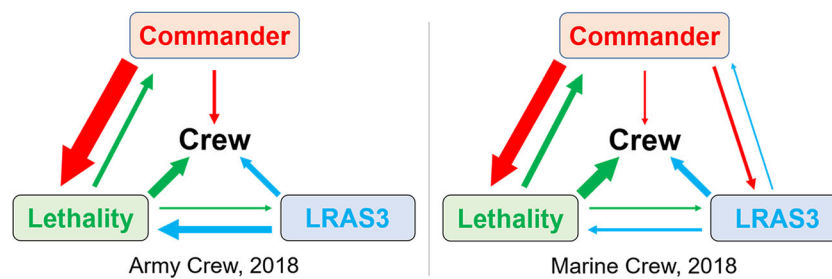
To produce aggregate communication flow models for a given team, the user first needs information about the sender and receiver of each communication event, as well as timestamps for each communication event. Communication events can be defined based on the communication modality utilized by the team and by the level of analytical detail needed. The operationalization for each of these characteristics rests with the researcher and the research question at hand. For a study using verbal interactions, a communication event might be defined as an uninterrupted phrase uttered by a single speaker, the "sender" defined as the speaker, the "receiver" as the intended listener, and the "timestamp" as the time that the speaker began to speak. For interactions through a chat messenger, the process of defining these terms is easier, as the needed data can usually be exported automatically. However, it should be noted that chat messages may not be read until a later time, so some consideration must be made as to what is used as a timestamp. Regardless, the needed data will usually take the form of an audio recording, an audio transcription, or an event log, such as an exported chat log from an instant message application. After preparing a dataset that contains the sender, receiver, and timestamp for all communication events, the user can then evaluate the number of

communication events accounted for by each sender-receiver pair across a given time period. Then, the user can create a flow map based on those proportions; using any graphic design software, the user represents each possible sender along with lines going from each possible sender to each possible receiver, with line widths corresponding to the proportion of communication accounted for by each sender-receiver pair. Some research efforts are underway to automate the process of generating flow maps and produce them real time based on team communication (Baker et al. 2020a, b).

This approach can be applied to teams of varying sizes and compositions, so long as there is information being sent and received between the teammates. The team members that send and receive data will populate the flow map, so this approach can also account for autonomous teammates if the autonomy is also sending and receiving communication. One extra challenge that can be experienced with human-autonomy teams, such as military teams, is the use of open communication channels such as intercom systems, where teammates monitor multiple channels and multiple possible receivers can hear, or not hear, a sender's message. This can make it more challenging to positively identify all recipients of a sender's communication.

One use case for this approach involved the Wingman manned-unmanned team (Brewer et al. 2018), which is composed of a robotic weaponized system and a single-manned command and control vehicle containing several human crew members. A dataset was collected from a Wingman gunnery crew, and clear differences were identified between the sender-receiver pairs exhibited during gunnery engagements versus during inter-engagement time (Baker et al. 2020a, b). Results using this approach showed that crews significantly restricted the diversity of who spoke to whom while on-task. In addition, the Soldier crew demonstrated a simpler, more rigid aggregate communication pattern consisting of fewer sender-receiver pairs, whereas the Marine crew exhibited a looser pattern characterized by more sender-receiver pairs (Fig. 1). Review of the transcribed data suggests that these patterns were related to the way that communication was used by the teams; the Soldier crew's communication was more task-related and direct, where the Marine crew's communication was more flexible and geared toward ensuring that all crew members maintained awareness of what was going on. Performance data from the gunnery event indicated that the Soldier crew scored higher on the gunnery qualification course, suggesting that the optimal communication configuration for this type of human-autonomy gunnery task may involve limiting the number of unique sender-receiver pairs.

This approach may be applicable to team cohesion. Because aggregate communication flow depicts the proportion of team communication accounted for by each pair of team members, team cohesion can be determined by following the strength of communication among team members over



**Fig. 1** Aggregate communication patterns exhibited by Army and Marine crews during a Wingman human-autonomy gunnery task. Arrows start at sender (speaker) and point to receiver (intended recipient).

The diagrams feature three crew member roles as well as “Crew,” which indicates that a communication event was not targeted at a specific crew member. Thicker lines indicate a greater proportion of communication

an extended time horizon (Hung and Gatica-Perez 2010). For example, aggregate communication flow can be used to identify time periods in which a team member becomes isolated from others (i.e., by communicating less often), and if this isolation is not due to specific task demands, an intervention can be implemented to avoid negative effects on team cohesion.

More work is needed to make the process of generating sender-receiving data more efficient, especially for verbal interactions. Automated transcription software is promising, but the latest software can struggle in environments with multiple speakers, background noise, or other suboptimal factors (Krausman et al. 2019). Further development in the detection and correct identification of senders (and recipients) will allow this approach to get closer to real time. Generating sender-receiver data real time from verbal communication will allow flow maps to be produced automatically for a given time period, allowing those observing a team (e.g., for training, test and evaluation, or research development) to see how teams are communicating at a composite level.

By design, flow maps are rapidly produced and easily interpretable, so aggregate communication flow is valuable in scenarios where immediate communication data is warranted. One such application is to military after-action reviews, which are key to supporting training and improving future operations in human-autonomy teams (Brewer et al. 2019). However, the strength of this approach begets two limitations. First, because it is an aggregate measure, it does not provide any information about changes in communication flow throughout the window of measurement, which can reveal stability, perturbations, or other dynamical properties (Cooke and Gorman 2009). Second, aggregate communication flow does not rely on deeper statistical analysis, and thus the insights provided are limited in scope. Other communication assessment approaches therefore expand on this by involving additional metrics to paint a clearer picture of the exchange of information within a team. Social network analysis (Section 2.1.2) takes the same data used in aggregate communication flow and applies additional metrics such as *centrality*. Following that, relational event models (Section 2.1.3) use similar data to identify particular structural

and temporal biases in the sequencing of communication over time. Then, we discuss anticipatory information pushing (Section 2.1.4), which also uses similar data to understand how preemptively sending information within a team relates to their trust and performance.

### 2.1.2 Social network analysis

Social network analysis (SNA) is an approach to understanding the relationships between actors in a network. It has been applied in a wide variety of fields including sociology, mathematics, and epidemiology and has also been used to analyze team performance (Balkundi and Harrison 2006; Borgatti et al. 2009; Shaw 1964; Walker et al. 2006). SNA provides a set of tools that may be useful for understanding the communication patterns that constrain and facilitate information flow and coordination in human-autonomy teams. Of particular interest is the group of measures related to *centrality* (Freeman 1978; Katz et al. 2004). Centrality generally describes the amount of connectedness which a team member has within a network. Measures of centrality may be useful for determining the prominence of an actor within a human-autonomy team’s communication process. For example, whether an IA or a human teammate has high individual centrality in a team’s communications can have implications for team coordination and performance. Humans may be more limited in resources for processing communications from multiple sources simultaneously (Wickens 2008) and high centrality may put them at risk of mental overload when communication volume or task complexity is high (Shaw 1964). However, a human’s capacity for integrating context and interpreting ambiguous information gives them potential for performing well in highly central roles in team-level decision-making tasks and leadership activities. IAs may not be limited in the same way as humans by their information processing rates and available channels, but exhibit brittleness in ambiguous or unanticipated situations (Endsley 2017) and be may less suited for highly central roles in tasks where they are likely to perform poorly. However, IAs may be well suited for highly central roles in which the task is straightforward and

requires data processing capacity (e.g., receiving and compiling messages containing resource quantities or teammate locations).

In addition to characterizing human-autonomy team communication and coordination, SNA may be useful for examining other aspects of teamwork like cohesion and trust. SNA can reveal aspects of the actual relations between teammates in contrast to what is formally delineated by organizational structure. For example, examining the centrality of actors in a network may reveal that the team leader is receiving an overwhelming quantity of communications during critical periods, suggesting that more initiative needs to be pushed to subordinate leaders. Examination of the direction and qualitative content of communications could also reveal communication imbalances or bottlenecks, such as an IA receiving a high volume of messages, but rarely passing it on to other teammates, or a change in communication networks following a breach of trust. Furthermore, SNA is not limited to analyzing explicit communications, but has historically been used in combination with survey-based approaches to measure subjective evaluations of other relational qualities like trust, influence, and affect. This can be used to generate other types of networks such as trust networks and teammate affinity networks (Borgatti et al. 2018). Such networks may provide additional insight into trust or cohesion measures (particularly between individual nodes) within the network.

A social network consists of a set of actors (“nodes”) and relationships (“links”) (Wasserman and Faust 1994). Links can be directed or undirected and can be assigned values and other attributes (Barrat et al. 2004). Entities in the network should be explicitly defined for the team communication dataset under consideration. For example, nodes may represent team members and links may represent explicit directional communications between agents weighted according to the number of transmission (e.g., Barth et al. 2015). Once network entities are defined, the team interactions can be analyzed. Captured data should, at a minimum, include message senders and receivers, but it may also be prudent to capture other aspects such as message length, content, and message times. Communications can be captured using voice recording software, text chat systems, or in some cases, estimates can be derived through observation or questionnaires, so this approach can accommodate the various communication modalities that may be used by an IA. Other elements such as the communication medium can also be useful for analyzing multimodal communications within a team. Once the team communications are recorded and formatted, measures can be computed to characterize the team’s social network. If such high-resolution communication data are not available, surveys about teammates’ interactions with fellow teammates can provide insight into the structure of relations within the team. Such survey methods lose the richness embedded in timing of interactions (e.g., enabling reconstruction of multiple

snapshots of the network over time) and are prone to cognitive biases in informant recall (Krackhardt 2014). Examples of social networking software include UCINET (Borgatti et al. 2002), statnet (Handcock et al. 2008), and igraph (Csardi and Nepusz 2006). SNA measures can be on either the individual actor-level or on the global network-level, and can be an aggregate of a single time span or split into epochs for comparisons over time (Borgatti et al. 2018). These respective measures may provide insight into how node-level, graph-level, or dynamic features of the network contribute to team trust and cohesion.

SNA can be applied to teams of various sizes from small to very large. One key area of research has focused on the relationship between various social network analysis measures of centralization and outcomes for individuals or groups. In highly centralized team communications, one or a few actors are involved in proportionally more communications than the rest of the team. Examining network-level measures like centralization of team communications can provide a team-level view of communication structure. Some research on the relationship between team centralization in all-human teams has suggested that less centralized structures may foster interdependence, coordination, rapid information sharing, and improved performance in complex tasks (Barth et al. 2015; Brown and Miller 2000; Shaw 1964), but other research has shown that high centrality of actors with access to resources may also improve performance (e.g., well-connected leaders; Sparrowe et al. 2001). Other research has suggested that highly centralized communications can result in communication bottlenecks which can cause errors and delays in information transfer between teammates and result in coordination deficiencies (Roberts et al. 2019).

Communication bottlenecks may play an important role in understanding *team workload*—a conceptual relationship between task demands, the team capacities, and task performance that emerges on the team-level (Funke et al. 2012). For example, communication bottlenecks may result in poor task reallocation and workload distribution. However, team communications that display a more decentralized structure and high network density may distribute information processing demands throughout a team and enhance the team’s overall workload capacity but increase individual workload in the form of communication overhead (MacMillan et al. 2004). Different communication network characteristics seem to be related with how a team performs in a specific task and under different levels of demand. For example, one study in command and control simulation found that both centralized hierarchical teams and decentralized peer-to-peer teams adopted communication patterns akin to a “small-world network” structure—a balance between local and global connectedness—when given the freedom to do so. This adaptation allowed for short distances between actors (e.g., access to information and resources) without the coordination costs of a fully connected network (Watts and Strogatz 1998;

Stanton et al. 2012). However, teams that fail to form adaptive communication structures in response to changing tasks may perform poorly in workload peaks. Thus, examining team communication network structures aggregated across specific events may help to characterize how they adapt their communications. SNA has also been used in combination with other human factors approaches to understand other team and system-level phenomena within sociotechnical systems such as distributed cognition (Plant and Stanton 2016) and distributed situation awareness (Stanton et al. 2006). However, SNA approaches have yet to be fully embraced for examining communications in human–autonomy teams.

SNA is an established methodology in many fields. However, it is not clear which findings from human-human teams will carry over into human-autonomy teams. Examining human-autonomy communications within various contexts will provide insight to the usefulness and generalizability of SNA. More work is needed to understand whether SNA measures relate to aspects of human-autonomy teamwork, such as trust and cohesion. One key consideration is how to treat the unique qualities of IAs in relation to humans within a team's network. For example, in human-human relations, trust is bidirectional (Grossman and Feitosa 2018). In human-autonomy teams, human teammates can certainly have trust in the IA (de Visser et al. 2019). However, trust does not reciprocate from the IA to the human, which could affect how SNA measures are collected or computed. Similar considerations should be made of communication-based data.

### 2.1.3 Relational event models

Although SNA has historically focused on descriptive measures of time-aggregated networks, relational event models provide a model-based approach for analyzing dynamic networks. A static representation of a network illustrates the structure of a set of relationships within an observed population; however, many different types of interactions underlie these relationships. Aggregation of interaction data obscures time-ordered information, edge weights, and changes of actor sets within a population (Quintane et al. 2014). By representing the network as a timestamped series of discrete, directed interactions called relational events, the relational event model preserves the richness of these interaction dynamics within networks (Butts 2008). This is a contrast to models of static, time-aggregated representations of the network. Because higher-order group constructs such as trust and team cohesion arise as a function of interaction dynamics (Marks et al. 2001; Schechter 2017), relational event models are well positioned to provide insight into the emergence and evolution of such states (Leenders et al. 2016).

The relational event nomenclature refers to a set of discrete actions in which one entity directs a social action to another entity or *set* of entities. This relational event framework can be

represented as  $a=(i, j, k, t)$ , where  $a$  represents a discrete event (i.e., an interaction),  $i$  is the sender of that event,  $j$  is the receiver of the event,  $k$  is the type of event (one may consider different types of interactions such as face-to-face communication, phone calls, emails, etc.), and  $t$  is the time of the event. Although  $a$  represents a single event, a series of relational events is represented in history  $A_t$ . The goal of the relational event model is to model this sequence of events  $A_t$ . Results from relational event models demonstrate interaction rates as a function of sender and/or receiver attributes, recent communication activity, and structural features of the network.

Data compatible with the relational event model are simple in structure: this method requires only a log of sender and receiver data. If exact timestamp data are unavailable, interactions can be modeled ordinally instead. Building on a framework for traditional event history models, the relational event model posits the joint likelihood of event history  $A_t$ ; that joint likelihood is simply the product of the likelihoods of each discrete relational event  $a$ . The conditional likelihood for the  $i$ th event in  $A_t$  is equal to the hazard for relational event  $a_i$  (following terminology used in the survival analysis subfield where “hazard” refers to the likelihood of a given event happening per unit of time). This hazard depends on a set of specified sufficient statistics, which may include sender attributes, receiver attributes, network statistics including the degree centrality, triadic closure, and past events that may be relevant to the given event (e.g., immediately reciprocating a communication act or the propensity of an individual to keep sending messages to the same partner). The model ultimately indicates how the relational event history (i.e., observed sequence of interactions) may be biased towards or against specific types of interactions.

There are several options for relating emergent states such as trust or cohesion to relational event model results. By using self-reported states as individual attributes, the model can be used to investigate directly how those states shape the propensity for an individual to send or receive interactions (e.g., someone who reports higher trust may send messages at higher rates, or discrepancies between individuals' ratings of group trust may be associated with fewer messages exchanged between them). By measuring rates of triadic closure, the model also allows us to evaluate the rate of structurally cohesive interactions. Finally, relational event model coefficients can be used to infer trust measures at the group level. In the latter approach, one would measure emergent states across a population of teams and run an identical relational event model on each team's communication. By relating those model coefficients to team-level states, links between interaction dynamics and team states can be identified (Pilny et al. 2014). Because team states emerge and evolve as a function of interactions, using features of those interactions to infer team states offers promising potential (Leenders et al. 2016).



Like many novel statistical methodologies, the growth in popularity of the relational event model depends, in part, on the availability of data compatible with the model. Although social network analysis has existed for nearly a century (Moreno and Jennings 1938), dynamic models are a recent development. As such, most social network data collected throughout the twentieth century reflected the field's then-limited methodological capabilities. Coincident with social network data collected explicitly in the relational event format over the last decade, similar behavioral logs have grown considerably in recent years due to the proliferation of electronic storage of cell phone records, email records, online interactions, and other electronic communication (Leenders et al. 2016). Scholars have used the relational event model to analyze radio communication networks during disaster response (Butts 2008), cooperative and hostile actions among nation states (Brandes et al. 2009), small teams playing cooperative games in laboratory environments (Leenders et al. 2016; Pilny et al. 2016; Schechter and Contractor 2016), intra-team interactions during a NATO exercise (Schechter 2017), interactions among college students over 6 months (Pilny et al. 2017), and individual contributions and bug fixes in a code repository for open-source software (Quintane et al. 2014).

Recent work on human-human teams working with an autonomous vehicle during a series of training exercises demonstrates how relational event modeling provides insight into team behavior in military contexts (Baker et al. 2020a, b). This paper used transcript data from a series of gunnery exercises to highlight distinctions between teams' message transmission patterns. Additionally, it demonstrated how centralization patterns diverge during dry fire and live fire exercises. As autonomy becomes increasingly integrated into these military teams, relational event models will help us to understand how the introduction of that autonomy shapes team communication dynamics.

Although relational event modeling has been valuable for understanding team interaction dynamics, greater availability of human-autonomy team interaction data will help us to determine how well our understanding of human-human team dynamics will translate to human-autonomy team dynamics. Data in which autonomy directly and naturalistically communicates with humans would be ideal (Demir et al. 2020), though more simplistic interactions (e.g., an IA sending an alert to a teammate) may also provide useful insight. Coupling those interaction data with high-resolution state dynamics will help us continue to understand how interaction dynamics shape team state dynamics.

#### 2.1.4 Anticipatory information pushing

We now transition from an approach that considers the timing and sequencing of interactions to one that considers flows across relationships. This approach considers a team

member's anticipatory information pushing (AIP) as a means to understand team trust, thus maintaining a focus on how team members send and receive information. In this case, information pushing involves information provided to a teammate after a request is made for it, whereas *anticipatory* information pushing specifically does not involve an explicit request. Therefore, AIP is a pre-emptive behavior in which a team member anticipates an information need of another teammate and provides that teammate with the necessary information they need before they ask for it. This is exemplified when "the right information (gets) to the right person in the right amount of time, in order to overcome an unexpected roadblock" (Demir et al. 2019; p. 146). Related concepts are found in transparency research, which has found that IAs can improve team trust, performance, and situation awareness by providing humans with information about the IAs' goals, plans, and decisions (Lakhmani et al. 2016; Stowers et al. 2016; Chen et al. 2018). For AIP, if information is passed preemptively, it might lead to higher levels of trust among teammates, especially in military settings, as the quality of the team's communication can mean the difference between life and death. Furthermore, because traditional methods of evaluating trust have been primarily static methods (e.g., questionnaires before or after a task, or interrupting a task), it may be possible to use AIP behaviors as an unobtrusive, continuous index for team trust.

Information pushing in human-autonomy teams has been measured in an unmanned aerial vehicle (UAV) synthetic task environment (Cooke and Shope 2004; Cooke et al. 2013; McNeese et al. 2018; Johnson et al. 2020a, b). This UAV task environment includes three teammates with heterogeneous and interdependent roles (pilot, photographer, and navigator) performing an aerial reconnaissance task under routine and degraded conditions. To perform well, the team must communicate with one another via text chat and safely navigate their simulated UAV along a series of waypoints while taking photographs of targets. Analyses of the communications in this task environment have suggested that pushing information is associated with better performance and team situation awareness (Demir et al. 2017). However, compared to all-human team, teams that had an IA as a pilot did not push information as effectively, suggesting a weakness in the IA's ability to effectively anticipate the information needs of teammates (McNeese et al. 2018). This approach to analyzing communication could gain further fidelity by focusing specifically on AIP, rather than information pushes in general. The text-based communication data can be coded for AIP by examining whether teammates pushed information to another teammate without first being prompted by the other teammates. The frequency of information pushing within the team (e.g., from an IA to a human, or from any team member to any other team member) can then be compared to how teammates self-reported their trust towards each other, which can be

accomplished with standard regression analysis procedures. Applying this approach to a given human-autonomy team can allow one to determine whether higher frequencies of AIP from a synthetic teammate to a human teammate correlate with higher levels of reported trust (Huang et al., 2020a, b, c).

This approach could also be used to contextualize or understand team performance. Research in the past has shown that teams that do more AIP perform better overall. A study conducted by Weber and Aha (2003) showed that in a military setting, teams that received information just in time during the plan authoring phase had overall better plan execution. Campion and colleagues (2012) looked at how information pushing in the context of health information exchange can result in a better overall experience for doctors in comparison to existing healthcare systems that rely only on information pull methods. In a study using a military context, Entin and Serfaty (1999) found that teams who were trained to change coordination behaviors when presented with high-stress situations experienced improved performance, with increased AIP from subordinates to leaders as well as better anticipation of teammates' information needs.

This approach can ostensibly be applied to any type of human-autonomy team that utilizes a communication modality that can be leveraged for AIP. In order for the results of this approach to be valid, there needs to be some variation in the reported levels of team trust, in order to determine how AIP behaviors fluctuate in the context of corresponding trust fluctuations. Otherwise, limited variation can lead to faulty data analysis. Another lesson learned from previously conducted research is that the pushing of information in human-autonomy teams may be less than in human-human teams (Demir et al. 2017). Therefore, future research is needed to shed light on how low levels of AIP in human-autonomy teams versus high levels of AIP in human-autonomy teams leads to differences in trust levels among human teammates towards autonomous teammates.

There are some limitations to this approach as there are not that many autonomous agents in existence that are capable of (a) being able to communicate with a human teammate and (b) able to replace a human teammate and do everything the replaced human teammate was responsible for doing. Therefore, measuring trust towards an autonomous teammate and correlating that trust with how much AIP occurs is currently limited to settings that involve autonomous agents with sufficient intelligence. Furthermore, communication is not always words (auditory/speaking or visual/text-chat communication). Sometimes it is symbolic, such as icons on a screen, so recognizing when an autonomous agent is doing AIP using symbolic communication can be difficult. However, it is clear that more research is needed to understand how and when autonomy should engage in AIP, as well as how the role of the autonomy might determine optimal AIP behaviors. Previous research (Cooke and Shope 2004; Cooke and Gorman 2009;

Cooke et al. 2012) has laid the foundation for how actual autonomy in a team setting would need to communicate with human teammates in order for the team to be an effective team overall. For example, as was mentioned earlier, autonomous agents have to be able to recognize the needs of human teammates and give them information in an appropriate amount of time based on those anticipated needs (Demir et al. 2017). Research designs that utilize different communication modalities will be especially relevant to this area, and as IAs become more advanced and capable of engaging in anticipatory information pushing, this approach will become more useful.

## 2.2 Dynamical systems approaches

The preceding four communication assessment methods have primarily focused on the structure and timing of team interaction. Other methods can evaluate team interactions with an emphasis on different timescales of the team's lifespan. The following two approaches build on this perspective by drawing from dynamical systems theory (Abraham and Shaw 1992; Gorman et al. 2017).

### 2.2.1 Distributed dynamic approach of team cognition

The theory of interactive team cognition (ITC; Cooke 2015; Cooke et al. 2013) suggests that team cognition is best measured at the team level through the analysis of the interactions among team components, including humans and non-human entities. The team interactions tie closely to the team composition, task context, interaction modality, and time scale. Traditional human-autonomy teaming research has focused on dyad relationships that involve one human and one IA. However, in many real-world applications, team components include not only one end-user and one IA but also multiple end-users and even other stakeholders (e.g., managers and engineers; Ho et al. 2017), as well as multiple IAs with either similar or different roles. The communication among these individuals and entities influences both human-human relationships and human-IA relationships. Moreover, the communications span over a long time frame as stages of an IA progress from concept and prototype development, testing and training, deployment, to retirement. Most literature has only covered the stage of concept and prototype development (Hancock et al. 2011; Hoff and Bashir 2015; Lee and See 2004; Schaefer et al. 2016), without considering the stages in the IA's life cycle where the technical readiness of the IA, the organizational acceptance of the IA, and the stakeholders' context of interaction with the IA all vary.

Because these issues will likely influence the generalizability and boundaries of research findings regarding human-autonomy communications, it is highly recommended to consider factors such as the stage of an IA, the team composition, interaction modality, and interaction duration in studies. In

addition, the distributed dynamic approach of team cognition aims to examine communication from a holistic perspective with an emphasis on distributed dynamic team cognition. Each stakeholder's cognition (e.g., trust) toward an IA does not only affect his or her own performance; it may also impact other stakeholders' cognition toward the IA, and even the stakeholder's attitude toward other stakeholders who share (or do not share) similar cognition. Interpersonal cognition and human-autonomy cognition mutually influence each other (Huang et al., 2020a). To understand these interconnected relationships among stakeholders and IAs, it is critical to study team cognition at different stages.

This distributed and dynamic team cognition approach is customizable and provides a holistic view of team effectiveness, as long as the threshold for the model components is defined. It can also be used to identify the source of issues in human-autonomy teaming because the approach investigates the team composition, interaction modality, and interdependencies in a network. The source of an issue could be a crew member or the IA, and could relate to the characteristics of the stakeholders or their functional tasks on the human-autonomy team. As an example, an issue could involve trust issues toward using an IA; this could be caused by the interpersonal relationship among the crew members, or by a specific task interaction that may warrant attention. An individual's interaction types and data may contribute to different aspects of the issues. For example, speech content type and flow patterns may indicate the interpersonal aspects of the trust. Increasing action duration and error rates may point to outcome-based trust issues.

The distributed dynamic approach of team cognition assesses communications in human-IA teaming while accounting for three aspects: (1) the stage of the IA's life cycle, (2) multiple stakeholders and IAs on a team, and (3) features of tasks and interactions among the individuals and entities. This approach has roughly five steps (Huang et al., 2020a):

**Step 1: Identify the area of interest and the topic of interest** It is important to identify the area of interest regarding human-IA teaming because context comes with many assumptions and rules as a background. One such example comes from the military context, the Next-Generation Combat Vehicle (NGCV) Cross Functional Team (CFT), which envisions manned and autonomous combat vehicles forming a team to support lethality of each unit as well as to improve the unit's safety and security. With the area of interest identified, the topic of interest (e.g., team trust, situation awareness, workload, resilience, and cohesion) will narrow the scope of relevant literature to review and subsequently inform which findings to replicate and which gaps to fill. Literature has shown limitations in its findings in human-human team trust and human-autonomy trust, such as its lack of consideration of stage, coverage of team components, and dynamic measures.

**Step 2: Identify the stakeholders and the artificial entities involved** This step defines the size of the team and its components. The analysis level of team size and components will determine the types of tasks and interactions to be included in the human-IA teaming. Communication with the sponsor may help determine priority interests. Interviews with subject matter experts are used to identify the relevant and key individuals on the chosen team level. For example, Ho et al.'s (2017) work illustrated a multi-stakeholder example involving the end-users (i.e., pilots), engineers (i.e., developers and technical personnel), and managers (i.e., trainers and high-level decision makers).

**Step 3: Analyze the communication modality and interactions among the individuals and entities through each dyad on the team** After further defining the subtopic of team trust as a phenomenon among interdependent team members that will help the team achieve their goals, the interdependent functional tasks need to be examined for each type of stakeholder. The interdependency and interaction analysis will produce an interaction taxonomy with communication modality (Huang et al., 2020a, b, c): (a) verbal communication—oral and text natural language through interviews, text messages, emails, technical testing reports, radio communication, documented daily conversations, internal forum posts, etc.; (b) visual interaction—gazing patterns; and (c) physical interactions—button pushes, hands on the wheel, tactile force, and so on. The communication modality depends on the time frame, types of stakeholders, and their tasks. Based on the communication modality and interaction taxonomy, collect interaction data that are accessible and analyze each type of data to cross-validate the findings (Huang et al., 2020b). This multi-method approach is recommended because the methods capture different aspects of complex team cognition and provide a more complete picture of the topic within the given context. For example, the three layers of trust (i.e., dispositional trust, learned trust, and situational trust; see Hoff and Bashir 2015) require different types of data to best fit each construct. Choosing the appropriate data and analysis method depends on both the literature and the accessibility of the data in the given context. A comparison of available methods using empirical data was developed to provide further guidance in this regard (Huang et al., 2020b).

Evaluation of verbal communication during a human-autonomy team operation should account for elements of the interaction (for a related interaction taxonomy, see Huang et al., 2020a, b, c). Using this taxonomy, the communication content theme and frequency can be coded to indicate whether the stakeholder trusts other individuals and entities on the team. There are two ways to code verbal data quantitatively and systematically: Coding Streams of Language (Geisler and Swarts 2019; Huang et al., 2020a) and Structural Topic Modeling (Lee and Kolodge 2018). A specific codebook for

the distributed dynamic approach is in development to identify the kinds of communication themes, frequency, and changes across phases for the context of Next-Generation Combat Vehicles and to compare the results with the two coding methods.

**Step 4: Plot the team analysis in a network** Communication data within a chosen period should be aggregated and plotted as a network of trust using nodes to indicate the involved team component and links to indicate trust relationships. The impact of managers', trainers', and engineers' trust toward the IA on the end-users' trust toward the IA can be noted on the network to account for the higher-level influencing factors that are beyond the dyad relationship between an end-user and an IA.

**Step 5: Analyze the trust network on a dynamic timeline** The communication-based trust network can be sampled and plotted along the timeline of an IA's life cycle or a specific stage. For example, in the stage of concept development, data samples can be chosen from the beginning phase of a mission, the perturbation phase, and the post-perturbation (i.e., recovering) phase. After we analyze sufficient interaction data samples as described in Step 3 and identify reliable interaction patterns for a team cognition topic (e.g., trust), real-time communication transcription services (e.g., Microsoft Azure cloud solution and Zoom auto transcription) may enable real-time detection of trust indicators. Comparing the trust network at different phases allows us to compare and contrast the team trust state at these phases and then determine intervening strategies.

Team trust is one example of applying this distributed and dynamic team cognition approach on a human-IA team (Huang et al., 2020a). The process can be customized to address alternative constructs. For team cohesion, these steps would be revised by defining team cohesion based on literature, determining team level, identifying cohesion-related stakeholders and relevant artificial entities, analyzing stakeholders' interdependent tasks and their interactions types and frequencies, developing a team cohesion codebook for the interactions, and finally showing them on a cohesion network along the timeline.

Critical to this model are the topics of team situation awareness, team workload, and team resilience, in addition to team trust. Team situation awareness may focus on the successful usage of critical information in their interactions. Team workload may be reflected in their response time or communication content with teammates. Team resilience may be broken down to which team component is unable to execute necessary interactions to recover from the perturbations on what task so that the whole team fails, or which team component is able to fill in extra functions through the interactions to cover another team component's functions so that the entire team succeeds. The key is to identify the essential features of the topic and

operationalize it through communication data and patterns. This approach can be applied to different teams by following the steps.

This distributed and dynamic team cognition approach is relatively newer and more sophisticated than traditional ways of studying team cognition. The approach could expand the previous research fields by including the interactions among multiple stakeholders and multiple autonomies, and the impact of a dyad relationship on the other dyads in a network. This approach applies to any type of team and is customizable in terms of team size, time frame, and research topics. The richness of interaction modality allows the approach to use the available types of interaction data to overcome the constraints of data availability in different cases.

A current limitation of this method is that there is a lack of an established database of interaction patterns for the topics of interest. Additionally, task analysis and interaction analysis are labor-intensive for identifying communication patterns and their relationship with target variables at the beginning of the process. There are three potential ways to overcome these limitations. First, reliable communication pattern standards and their connections with the target team variables should be investigated through more empirical studies, so that it will be easier for future studies to use the pattern options and apply them to other topics in other contexts. Second, context-free communication patterns should be identified. For example, the duration of communication in different scenarios can be measured through a talk pedometer called LENA technology, which was initially used to measure children's talking volume without dealing with the contents (Wang et al. 2017; Odean et al. 2015). The context-free measures could reduce the amount of labor required to analyze all the communication contents. In the last decade, researchers have been using dynamics techniques to analyze patterns of content-free social science data (Amazeen 2018; Gorman et al. 2010). The dynamics analysis techniques allow for analyzing many interaction data types that traditional surveys and statistics cannot do. For example, the damping model could describe patients' improving accuracy of pain estimation (Finan et al. 2010) and may apply to users' improving trust calibration. Finally, automated transcription techniques, real-time data analysis, and automated data visualization may further improve the efficiency of this approach for researchers and practitioners.

### 2.2.2 Quantifying exploratory communication

This section expands on using a dynamical systems approach to team communication by providing a means to quantify how teams find new ways of coordinating and communicating. Whereas behavior may "exploit" previously effective solutions, behavior may also "explore" solutions which may potentially be effective. Exploration and exploitation tradeoffs have been researched across several domains including

machine learning (Kaelbling et al. 1996), animal foraging (Cook et al. 2013), and cognitive systems (Hills et al. 2015). For instance, work by Rolf et al. (2011) showed that efficient motor learning in infants could be accounted for by treating exploratory movements as goal-directed, rather than random. Their key insight was that feedback from exploration could be rapidly exploited to further approximate a target movement. Team communication may also be thought of in terms of exploration and exploitation, in which new ways of coordinating are discovered by varying communication to meet shared goals. Therefore, this method focuses on quantifying novel or *exploratory communication* to identify those patterns in human-autonomy teams. Although exploratory communication has not been explicitly defined, there are numerous definitions of exploration throughout the literature, whereby exploratory communication may be defined as communication with properties unique to a collective's interaction history (Hills et al. 2015).

Newly formed human-autonomy teams must learn to work together, which means identifying communication that works for that team. This process involves exploring communication to develop trust, cohesion, and accomplish team-level goals. For instance, a human teammate may be able to verbalize commands to an intelligent agent or query them for more information. By exploring communication with that agent, the human teammate may learn that agent's boundaries for appropriate reliance. Conversely, if a teammate explores and that exploration is attributed as a communication error, trust in that teammate may decrease. Teammates that deliberately explore together, perhaps indicated by attractor reconstruction (e.g., Gorman et al. 2010), may also be more cohesive, whereas more unstable exploration may indicate a lack of cohesion. Finally, because exploration is typically motivated by an intent to meet a shared goal, patterns should correspond generally with task novelty. A high-performing team may learn the appropriate communication processes necessary for meeting these goals more quickly. The communication patterns associated with efficient learning are likely to be non-linear, making non-linear dynamical systems methods appropriate.

Because teams may be required to communicate in new ways to adapt to novel or challenging circumstances, exploratory communication may be directly indicative of team-level adaptation and resilience. There is evidence that teams which have "metastable" coordination, or coordination that shifts between stable and unstable patterns with ease, are highly adaptive in the face of roadblocks (Demir et al. 2018). The first approach described in this section treats variability as an index of exploration and may be applied to measure the flexibility of a team's communication pattern in general, whereas the second approach addresses patterns of exploration specifically.

Non-linear dynamical systems analyses assume that events over time are sampled from a dynamical system which may only be in one state at a time. Therefore, communication

should be measured over time and interpolated as needed to produce a time series with a consistent sampling rate (e.g., 100Hz) and a defined set of possible states. A phase space, or a set of possible trajectories for the team's communication, will need to be constructed if the measured communication is not a continuous numeric variable. This can be done by decomposing the signal into a set of dimensions and states corresponding to each possible combination. Because exploratory communication generally refers to goal-directed communication variability, it is imperative that the aspects of communication measured are most relevant for the task or variable of interest. Although these may be specific to the task, some examples include communication flow, communication modality, current task, and communication content. Note that the set of possible dimensions such as communication content may need to be condensed into more tractable units such as the information conveyed or themes. For a detailed example of this process applied to team communications data, see Gorman et al. (2019).

Next, the team's behavior may be captured by using attractor reconstruction. Attractor reconstruction involves firstly estimating the length of a set of related events in the time series,  $\tau$  which is indicated by either the first point in time that is minimally correlated with the series or by calculating the average mutual information for that series. Then, an embedding dimension must be determined by identifying the number of dimensions that minimizes the percent of false nearest neighbors in the phase space. If the embedding dimension is less than or equal to three, then the phase space may be plotted and visualized. Attractor reconstruction is typically followed by calculating the largest Lyapunov Exponent to determine stability and recurrence quantification analysis to describe the predictability of a team's overall coordination pattern, represented as a recurrence rate and percentage of determinism. In addition to attractor reconstruction, several other methods for describing variability within a time series are available (Amazeen 2018), many of which show promise in the human-autonomy team context.

Another approach is to use qualitative coding to identify exploratory communication. For this method, it is critical to identify what exactly counts as exploratory. In controlled experiments, teams are often formed from individuals who do not know each other, and even familiar teams may lack experience working together on the specific task. Therefore, it may be safe to assume that initial team communications are exploratory in some experimental contexts. For experienced teams performing familiar tasks (e.g., a military squad that has trained together), some investigation into the team's routine communication practices may be needed to scope communication that has already been explored. One approach may be to assume a trained team as familiar with pre-defined essential coordination and to treat non-essential coordination as exploratory. Once data have been coded and translated into a set of

relevant states, dynamical methods may be applied. The difference between the first approach and this one is that this approach operationalizes exploratory communication as the signal itself, rather than the variability of an overall communication signal.

A frame of reference is critical for making sense of communication dynamics analyses. Although a communication pattern may be different (i.e., more or less stable) than another, this information does not address which pattern is better for that specific task or better for improving a variable associated with team effectiveness. In that sense, dynamical methods are excellent for describing a pattern and taking into account differences in communication over time which traditional static measures do not, but the association between variables such as trust and cohesion must be developed and validated with solid grounding in the context of teamwork. All teams are likely to exhibit exploratory communication, particularly when they are unfamiliar. However, exploratory communication is constrained primarily by the redundancy and degrees of freedom offered by the communication modality (e.g., natural language) used as well as the team interdependencies from which coordination is likely to emerge. Therefore, the proposed method is most appropriate for dynamic and complex tasks for which there are many possible solutions to achieving team-level goals. The teamwork context should contain enough variation to motivate exploration in teammates' behavior, and that variation should have impact at the team-level rather than the individual-level.

There is some evidence that different training regimes such as perturbation training (Gorman et al. 2010) which constrain team coordination facilitate exploration and adaptive performance. Additionally, research showed that mixing up team composition after a retention interval had similar effects (Gorman and Cooke 2011). In general, it appears that diversifying the constraints in teamwork may be a critical component to preparing teams for unexpected conditions to the extent that it affects how team coordination evolves. More development is needed to understand more broadly the patterns of exploratory communication associated with team effectiveness as well as how to instantiate those patterns through work systems design.

Notably, IAs do not explore communication in the same ways as humans, which may influence human-autonomy teams to coordinate more rigidly (Demir et al. 2018). In the short term, intelligent agents are unlikely to be able to explore communication in a goal-directed way. They will typically rely on fixed programming or training data to coordinate in a team. Future intelligent agents may have machine learning capabilities that allow for their communication to evolve over time. Measuring exploratory communication in teams with these agents may be useful for assessing the performance of these algorithms. Given progress in modeling exploratory communication, machine learning capabilities may benefit from extending the explore-and-exploit tradeoff to the communication behaviors of intelligent agent teammates.

## 2.3 Emotional states

The preceding assessment approaches have all shared a focus on the sequencing and interaction patterns of team communication events. These approaches rely on data that preserve the sequence of team communication events over time, such as transcribed audio, text or chat logs, or event logs, such as those exported from simulator systems. However, other assessment methods can leverage *features* of the team's interactions to provide insights into team states such as trust or cohesion. The next section describes an approach for performing emotional feature processing using facial features, followed by an approach for detecting emotional states using vocal features.

### 2.3.1 Facial expression analysis

One of the strongest indicators for emotions is the human face. We can read emotions in others based on changes in key facial features such as eyes, brows, lids, nostrils, and lips. The human face includes over 40 structurally and functionally autonomous muscles, each of which can be triggered independently of each other (but are likewise innervated by a single nerve, therefore referred to as the facial nerve). The facial nerve emerges from deep within the brainstem and branches off to all muscles like a tree. Here, facial muscle activity is highly specialized for expression—it allows us to share social information with others and communicate both verbally and non-verbally. Facial expressions are only one of many indicators of human emotion, but they might be the most apparent ones. Humans can produce thousands of variations; however, there is only a small set of distinct facial configurations associated with certain emotions, irrespective of gender, age, cultural background, and socialization history (to an extent). These are joy, anger, surprise, fear, contempt, sadness, and disgust.

Computer-based facial expression analysis attempts to mimic our human coding skills as it captures raw, unfiltered emotional response towards any type of emotionally engaging content. As such, emotional feature processing involves the detection of human emotional states from specific features of the face (e.g., specific movements of the face which reveal changes in universal emotions). Facial expressivity has been shown to be related to appraisal and coping mechanisms, as well as stress, fatigue, and trust. Past studies have found that automatic computations of facial expressivity are comparable to manual annotations of emotional expressions (Neubauer et al. 2017) and have been utilized in a number of both clinical and experimental studies (DeVault et al. 2014; Scherer et al. 2016; Venek et al., 2016; Parra et al. 2017; Batrinca et al. 2013; Chollet et al. 2015). Therefore, the platform provides evidence that automatic behavior trackers have the ability to support clinical assessments and provide researchers with much needed objective assessments of behavioral indicators of stress, trust, or even team cohesion, though to date, research

into these possibilities has been largely 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 three different ways. First, tracking of facial electromyography (fEMG) records the activity of facial muscles with electrodes attached to the skin surface. fEMG detects and amplifies the electrical impulse generated by the respective muscle fibers during contraction. For example, the corrugator supercillii (e.g., eyebrow wrinkler) is a small, narrow, pyramidal muscle near the eyebrow, generally associated with frowning. The corrugator draws the eyebrow downward and towards the face center, producing a vertical wrinkling of the forehead. This muscle group is active to prevent high sun glare or when expressing negative emotions such as suffering. In addition, the zygomaticus is a muscle that extends from each cheekbone to the corners of the mouth and draws the angle of the mouth up and out, typically associated with smiling. Therefore, when facial expressions are apparent, so too are the associated electrical impulses that result from movement.

The second method involves live observation and manual coding of facial activity using the Facial Action Coding System (FACS). The FACS represents a standardized classification system of facial expressions for expert human coders based on anatomic features. Coders examine videos of an individual's face and describe any occurrence of facial expressions as combinations of elementary components called Action units (AUs). Each AU corresponds to an individual face muscle or muscle group and is identified by a number (AU1, AU2, etc.). All facial expressions can be broken down into their constituent AUs. To use an analogy, facial expressions can be likened to "words," while AUs are the "letters" that make up those words. Table 2 illustrates which AUs can be calculated to reveal changes in universal emotions. For

example, the universal emotion "anger" is composed of evidence from AUs 4, 5, 7, and 23, which together can be averaged to reveal the overall evidence of a specific emotion, within an individual frame of data.

The third (and quickest method) utilizes computer vision algorithms to automatically detect a human face and employ feature detection to detect facial landmarks such as eyes and eye corners, brows, mouth corners, and the nose tip. With the feature detection, an internal face model is adjusted in position, size, and scale to match the respondent's actual face, as if an invisible virtual mesh was put onto the face of the respondent. Whenever the respondent's face moves or changes expressions, the face model adapts and follows. The feature classification then translates the landmark facial features into action unit codes, emotional states, and other affective metrics.

The individual's facial expressivity can be quantified automatically through several open-source or paid software licenses. For example, some commercial software (e.g., iMotion's Facet) provides automatic extractions for emotions such as *anger*, *sadness*, *joy*, and *contempt* on a frame by frame basis. Additionally, the OpenFace software platform (Baltrušaitis et al. 2018) provides automatic assessments of single action unit (AU) evidence, which can then be used to calculate universal emotions following the FACS calculations mentioned above (Ekman and Friesen 1978). These two technologies are widely used in the affective computing literature; however, they require offline analysis (i.e., video of an individual's face is processed after they engaged in a task). As such, there are additional technologies that allow for real-time analysis of facial expressivity (e.g., Visage Technologies). For all software systems, participants' facial expressions should be continuously recorded via a webcam embedded in their monitor or mounted atop their computer screen while they are engaging in a task of interest (e.g., a team coordination task). We note that the software packages discussed thus far are not meant to be exhaustive, but merely provide examples of the software that can be utilized for these analyses.

Many metrics to measure operator state exist and typically include questionnaire assessment; however, these are taken after an operator performs a task, requiring them to remember how they felt in a given moment, and possibly reflecting subjective biases. Additionally, unimodal streams of data may not accurately capture all aspects of an affective state or decision. In this context, it is vital that IAs not only accurately perceive human affective states but also respond appropriately to avoid misinterpreting social cues during collaboration to improve decision-making and performance (Scheutz et al. 2006). Most of the published research on computer vision approaches to operator state detection have focused on fatigue assessment and typically relied on analyses focused on eye tracking and head movements (Dong et al. 2011; Gu and Ji 2004; Zhang and Zhang 2006). In contrast, the relationship between team

**Table 2** Facial expression emotion calculation from single AUs (Ekman and Friesen 1978)

Emotion classification	Action units that comprise an emotion
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

AUs correspond to individual muscles or groups of muscles which can be used to calculate and categorize universal emotions

cohesion and facial expressivity has not been studied thoroughly. Therefore, we posit that these methods for measuring emotional response will provide more directed insight on understanding affect-based trust and team cohesion. This line of research is critical because it will be necessary to develop autonomy-enabled systems that can robustly perceive and respond to our emotions as we interact with them if human-autonomy teams are to be successful (Bartlett et al. 2004).

Through a large review of the literature, it has been found that there are six identified types of trust that impact human-autonomy teams: *trust propensity*, *trustworthiness*, *affect-based trust*, *cognitive-based trust*, *situational trust*, and *learned trust* (Schaefer et al. 2020). Facial data evaluation provides an additional means to evaluate *affect-based trust* within human-autonomy teams. Affect-based trust is an emergent attitudinal state in which the individual makes attributions about the motives of the automation (McAllister 1995; Burke et al. 2007). 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 which have been linked to affiliation, empathy, and assessments of team member trustworthiness.

Human-autonomy teaming presents an interesting case because human-human teams may communicate non-verbally through changes in emotional expression (i.e., constant search for information from our partners' faces). For example, if an individual worries about a particular decision they made or need to make, they may seek confirmation or alternate solutions from non-verbal features of their partner. Alternatively, if something negative impacted the state of the team and one team member responds appropriately (e.g., some sort of negative affective response) whereas the other does not (e.g., they smile in response to a team failure), then trust and eventual cohesion may suffer. Within the human-autonomy teaming domain, it is important to acknowledge that humans may not get the same kind of non-verbal feedback that they normally would from a human equivalent. Given scenarios such as these, additional considerations should be made with regard to the communicative design of human-agent teams.

### 2.3.2 Vocal feature assessment using neural networks

Individual emotional regulation and sensitivity to the emotional state of others on a team can be helpful when establishing team trust and cohesion. Most humans do this to some degree by interpreting facial expressions and vocal content in context. Good teamwork often consists of knowing when to pass along information and knowing when this information will not be helpful due to the constraints of the recipient's state (Lingard et al. 2004). Although the content of speech may be useful for detecting emotional state, this is most often communicated through vocal features such as the spectral and temporal characteristics of speech relative to one's neutral

speech. Stress can cause the muscles in the body to tighten, and this extends to the chest, throat, neck, jaw, and vocal cords (Hansen and Patil 2007).

Recently, neural network models have been shown to be able to effectively detect emotional state from the acoustic features of speech (Casale et al. 2008; Koolagudi and Rao 2012; Stuhlsatz et al. 2011). This capability has utility for teams consisting of humans and intelligent agents, as this allows adaptive autonomous assistance in response to stress and work overload. This capability has already been used for customer service applications like automated telephone attendants as well as for aid in psychiatric diagnoses of depression and post-traumatic stress disorder (Banerjee et al. 2017; Cannizzaro et al. 2004; Vergyri et al. 2015; Vidrascu and Devillers 2005; Vogt et al. 2008; Yacoub et al. 2003).

The ability to adapt one's behavior according to the needs of the team would seem to correspond to the trust and cohesion within a team and, by extension, to increased team performance. However, it has not been proven that improving the ability to detect and adapt behavior to emotional states will lead to increases in team performance or team outcomes. Thus, a first step is to establish a correlation between positive emotional states with other measures of trust and cohesion. It is presumed, but also unconfirmed, that positive emotional states will also correspond to improved team performance and, conversely, that poorly performing teams will have greater rates of frustration and negative emotions. A future goal is to identify when the human members of a team are experiencing a higher workload so that IAs can implement adaptive assistance. It is hypothesized that team trust, cohesion, and performance (as assessed by other measures) will improve as a function of the implementation of adaptive autonomous aids. Because the development of technical adaptations is beyond the scope of this work, the initial goal here is to simply establish a correlation between measures of emotional state and team trust and cohesion.

Neural network models are derived by training a model on a labeled set of data, in this case digital auditory data (e.g., .wav files) recorded from speech. Typically, the first step is to extract, as input, acoustic features from the speech signal. These features usually include information about the spectral content, the log-energy content, the pitch period, and statistical information about their averages, minima, and maxima. Derivatives of these features provide information about feature changes over time (El Ayadi et al. 2011; Schuller et al. 2003). The task of the algorithm then is to estimate the parameter weights of these features with respect to the probability of an outcome, and by optimizing these weights, the algorithm is trained to accurately classify the speech input into categories of emotional state. Models can be optimized using mathematical techniques, such as the use of hidden nodes, convolution, and sequential features like recurrent bidirectional networks and long short-term memory. These allow the model to



incorporate contextual information about speech, adjusting parameters based on what occurs directly prior to, and after the current speech content. The processing power required for such models depends on the model, but many algorithms can be implemented on standard CPUs (computer processing units) and GPUs (graphic processing units). Implementing the model requires access to a set of labeled speech files and time for the initial model training. Real-time implementation would require the use of a trained model, recording capabilities, processing capabilities that enable the extraction of features, and continuous reading of ongoing speech. Currently, such models are able to correctly identify 4–7 emotion categories at rates of 70% or more.

This method has been demonstrated to work and has been used for multiple applications such as detecting severe depression (Cannizzaro et al. 2004) and interactive voice systems (Yacoub et al. 2003). However, there are still some technical challenges to implementation of this model. First, there is the acquisition and labeling of a dataset of speech recordings for use in model training. Although pre-labeled sets exist, most are small (Bou-Ghazale and Hansen 2000; Burkhardt et al. 2005; Fiscus et al. 1993; Swain et al. 2018; Ververidis and Kotropoulos 2003; Ververidis and Kotropoulos 2006). Most existing sets use “acted” emotional speech (Liu et al. 2018; Vogt and André 2005). In real-world contexts, most speech is neutral, so it becomes the task of the model to identify anomalies, of which there are few to train on. Furthermore, most of the studies conducted have been in quiet environments with only a few exceptions (Huang et al. 2019). A potential complication for initiating this methodology for human-autonomy teams is that the environments are noisy. Finally, most models are somewhat accurate for the detection of multiple emotional states, but they have been most successful at discriminating between positive and negative emotions (Casale et al. 2008). In the context of team performance, this may be sufficient to enable IAs to detect when aid is needed.

The model currently under development for use in human-autonomy teams is a convolutional recurrent neural network model that incorporates long short-term memory and an attention layer. It is loosely based on the model proposed by Huang and Narayanan (2017). The model is trained on noisy, field study audio-recordings and, rather than learning multiple emotional states, and is trained to identify only stress states. The current implementation of the model has been trained on an existing speech dataset called the IEMOCAP dataset (Busso et al. 2008). Current efforts include the development of a labeled dataset from recent US Army field studies using the Wingman human-autonomy lethality platform in a gunnery task (Schaefer et al. 2019a; Schaefer et al. 2019b).

After training and validation, future efforts will include the development of a continuously operating model for real-time analysis of team performance.

## 2.4 Linguistic synchrony

The preceding two sections have illustrated how features of vocal and facial expressions could be leveraged to understand team trust and performance. These features can be captured real time, so while the approaches are in development, they are highly promising as opportunistically sensed metrics. Whereas the preceding two approaches relied on behavioral aspects of communication, the approaches described in this section utilize the *content* of communication to produce insights.

In teams, members must form a common understanding of their goals, roles, and procedures (Klein et al. 2005). To build this understanding, members must use strategies to transmit and encode relevant information (Wilson and Sperber 2012). In doing so, people engaged in a conversation (speakers) may exhibit similar lexical (i.e., word choice) and syntactic (i.e., sentence structure) properties in their utterances, presumably because greater lexical and syntactic alignment improves efficiency in communication (Semin 2007). In fact, many scholars have examined linguistic similarity among speakers, revealing positive associations with cohesion (Dong 2005; Heuer et al. 2020), trust (Scissors et al. 2009), and task performance (Dong et al. 2004; Foltz et al. 2003; Fusaroli et al. 2012; Gorman et al. 2003; Richardson et al. 2019; Yilmaz 2016). Here, we outline approaches which have yielded some promising results in human-human teams, including their computation, relevant findings in the literature, and how they may be adapted and improved for understanding cohesion, trust, and performance in human-autonomy teams. The linguistic synchrony metrics detailed here are designed to reflect different states and processes pertaining to communicating and encoding knowledge. Described here are three metrics: Language Style Matching (LSM: Niederhoffer and Pennebaker 2002), Latent Semantic Similarity (LSS: Landauer and Dumais 1997), and Conversation Level Syntax Similarity Metric (CASSIM: Boghrati et al. 2018).

LSM can be viewed as an analysis approach which focuses on the rate at which speakers use content-free words (i.e., words which, by themselves, do not carry any semantic meaning). The LSM approach assumes that explicit information becomes burdensome when communicating about a topic at length, and therefore speakers will tend to omit explicit details once common ground has been established, resulting in greater communication efficiency (Gonzales et al. 2010). LSM relies on the proportion of function words used in a conversation—the greater the similarity between speakers in their use of function words, the greater degree of linguistic synchrony. Function words include nine lexical categories:

articles, adverbs, conjunctions, negations, impersonal pronouns, personal pronouns, prepositions, verbs, and quantifiers (Gonzales et al. 2010). To compute LSM between speakers, the following formula is used:

$$\text{LSM} = \frac{\sum_{i=1}^9 \left( 1 - \frac{|p_{is_1} - p_{is_2}|}{p_{is_1} + p_{is_2}} \right)}{9}$$

where  $i$  represents each of nine function word categories (see above),  $s_1$  and  $s_2$  represent each speaker for the dyadic comparison, and  $p_{is_n}$  represents the proportion of function words within a category used by the speaker. Note that the difference between proportions is absolute, that LSM represents an unweighted average across the word categories, and that the possible values of LSM range from 0 (absolutely no matching) to 1 (perfect matching). Although the computation itself is simple and intuitive, it relies on the use of a reference—a dictionary—to accurately tag each word with its associated category, which has been made easy by tools such as the Linguistic Inquiry and Word Count software package (LIWC: Pennebaker et al. 2001).

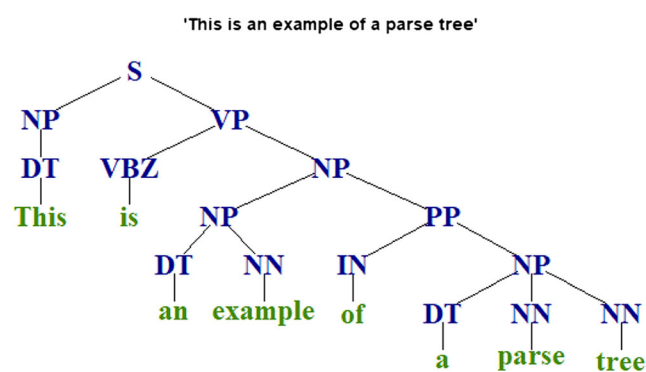
In contrast, LSS can be viewed as an analysis approach which focuses on the explicit details of a conversation, and that the semantic coherence between speakers' utterances, or between a speaker's utterance and known topics, reflects coherence and shared comprehension (Landauer and Dumais 1997). Rather than rely on pre-specified word categories, LSS relies on the use of Latent Semantic Analysis (LSA: Landauer et al. 1998) to construct word meanings based on their co-occurrences with other words. First, conversation transcripts need to be formatted in a matrix, with each column representing a document (e.g., an utterance), each row representing a term (e.g., a word or bigram), and each cell representing the number of times the term appears in a corresponding document. Using this matrix, LSA applies Singular Value Decomposition to reduce the dimensionality of the matrix, a method akin to principle components analysis. At this stage, the analyst must choose the number of topics (or factors) to extract from this matrix, which may be informed by several criteria that will not be explored here in detail; we instead refer the reader to existing resources (e.g., Evangelopoulos et al. 2012; Landauer et al. 2013). Once factorization is performed on the original matrix, one can then reconstruct the original document-term matrix, with cell values representing each term's *relevance* to each document in semantic space. The consequence of this transformation is that terms which may have never appeared in a document will, nevertheless, reveal themselves as *relevant* to a document by virtue of their associations with other terms which do appear in the document. Finally, the analyst can correlate, via cosine similarity, terms and documents with each other—if documents represent unique utterances, the cosine similarity between utterances represents their coherence in semantic

space, or their Latent Semantic Similarity. Because LSS is based on correlations, values can range from  $-1$  (anti-synchrony) to  $1$  (perfect synchrony).

CASSIM, the third approach, can be viewed as an analysis approach which focuses on the format in which information is communicated (Boghrati et al. 2018); several reports indicate that semantic and syntactic information is processed differently (Dapretto and Bookheimer 1999; Hagoort 2003), and that syntactic variation has meaningful implications for comprehension (Bock 1982). Below, we introduce the methods underlying each of these linguistic synchrony metrics. CASSIM is a fully automated process of estimating syntactic similarity (Boghrati et al. 2018) that directly analyzes syntactic structures by extracting constituency parse trees (see Fig. 2 for an example), which depict the breakdown of a sentence (S) into its substructures: noun phrases (NP), verb phrases (VP), and prepositional phrases (PP), which themselves comprise specific word categories, such as determiners (DT), third-person singular present verbs (VBZ), nouns (NN), and prepositions (IN).

With the structure of the sentence represented as a graph, one sentence can then be compared to the structure of other sentences, which CASSIM accomplishes using the edit distance, an algorithm that computes the number of changes required to convert one sentence structure to another—the fewer edits required, the more similar the syntactic structure. The changes can take three possible forms: insertions, deletions, and renaming. After extracting constituency parse trees and computing edit distances between sentences, CASSIM subtracts distance scores by 1 so that larger values (1 is maximum) represent greater similarity, while lower values (0 is minimum) represent lower similarity. If computing similarity at the level of an entire conversation, CASSIM employs the Hungarian Algorithm to find optimal pairings of sentences for computing similarity.

Several accounts have focused on how linguistic synchrony functions to build interpersonal trust, cohesion, and performance. For instance, LSM has been linked to greater social support in teams (Heuer et al. 2020), greater cohesion



**Fig. 2** Constituency-based parse tree visualization, which shows the nested structure of a sentence, comprising each of its syntactic components

(Gonzales et al. 2010), and better performance (Gonzales et al. 2010; Yilmaz 2016). There is also evidence from research on linguistic entrainment, or the use of similar words at similar rates between speakers, which suggests that entrainment on high frequency words (similar in computation to LSM: Rahimi, Kumar, Litman, Paletz, & Yu, 2017) corresponds to greater performance (Friedberg et al. 2012; Nenkova et al. 2008). Similarly, LSS has been used to predict positive social dynamics, such as greater interpersonal attentiveness (Babcock et al. 2014), and greater team performance (Dong et al. 2004; Foltz et al. 2003; Gorman et al. 2003; Martin and Foltz 2004). Although CASSIM has yet to be evaluated as a predictor of team performance, initial results demonstrate its ability to discriminate between related and unrelated text responses (Boghrati et al. 2018), and that subordinates will adjust their syntax to accommodate superiors (Boghrati and Dehghani 2018).

Although these linguistic synchrony metrics are typically used to understand dyadic communication, researchers also apply these methods to multi-party dialogues, which they accomplish by averaging across dyadic contributions to team communications (Litman et al. 2016; Rahimi et al. 2017), or comparing each member to the group as a whole (Gonzales et al. 2010). There are no known limitations on team size, structure, or composition with respect to using these metrics; however, researchers ought to be vigilant about the constraints of their own study design when interpreting the meaning and relevance of these linguistic synchrony metrics. For instance, analyses could be conducted across teams, across members within teams, or within members over time. Such flexibility will result in several distinct models depending on the dimensions of interest and any contextual information available (e.g., roles, hierarchy, task constraints). As a specific example, Yu et al. (2019) used LSM to explain variation in the degree to which team members self-reported experiencing conflict during a cooperative board game, the results of which indicated that teams who experienced greater change in their degree of LSM also self-reported less conflict. By contrast, Gonzales et al. (2010) used a static measure of LSM, rather than change in LSM across the task, to characterize communications during a group coordination task, the results of which indicated that teams with higher LSM performed better on the task. These examples illustrate a small divergence compared to other uses of these same underlying linguistic synchrony metrics, especially when considering how these metrics become incorporated in larger causal models.

Beyond the flexibility with which these metrics can be analyzed, there are inconsistencies in understanding the importance of linguistic synchrony in team processes. For instance, some scholars have found that greater LSM is associated with worse performance (Heuer et al. 2020), is not associated with performance at all (Munson et al. 2014), and that its associations with performance are moderated by other team

features (Gonzales et al. 2010; Yilmaz 2016). Drawing strong conclusions regarding the role of linguistic synchrony in team performance is difficult at this time, because very little evidence was obtained using the same methods for collecting and analyzing data. However, automated transcription technologies are becoming better and more widely available, reducing the burden of analyzing naturalistic communication data. Therefore, an influx of communication data from research on teams will help clarify the relationship between linguistic synchrony and team performance.

In human-autonomy teams, the applicability of these methods will depend on the role of an IA, which can take two forms: as communication observer and as communication participant. An observing autonomy would be poised to assess linguistic similarity in real time and provide feedback to those with vested interests. However, a participating autonomy would need more sophisticated capabilities, such as understanding how these linguistic features correspond to team processes and the external environment, or understanding how to produce natural language to make lexical, semantic, and syntactic features comparable across entities. The dependence on natural language is particular to LSM, which relies on matching words to a specific dictionary, and CASSIM, which appears to be only reasonable for comparing syntax within a modality (e.g., natural language to natural language). On the other hand, LSS relies on the association between generically defined “terms” and “documents,” making it amenable to understanding the semantic relevance of behaviors (Chen et al. 2019; Niebles et al. 2008), which could include any non-linguistic communication.

As our ability to capture accurate real-time communication data becomes more tenable, collecting large corpora of human-autonomy team communication data will enable more rigorous testing of these metrics. Even with our current analytic capabilities, it is worth evaluating the performative difference between completely accurate transcriptions and fully automated speech recognition transcriptions; in fact, early research demonstrated only a 10% reduction in predicting performance when using automated transcriptions with a 57% error rate (Foltz et al. 2006).

As more capable conversational agents are developed, these metrics can inform autonomy about how to actively engage in the grounding process. To achieve this, autonomy must be able to incorporate information beyond transcripts, such as where people are looking (Altmann and Kamide 2007; Knoeferle and Kreysa 2012; Staudte et al. 2014) or gesturing (Beilock and Goldin-Meadow 2010; Galati and Brennan 2014; Goodwin 1986). For autonomy to be fully participatory, it must also have the capacity to learn to adapt its own communication to its human teammates to ensure optimal efficiency and efficacy of bidirectional human-autonomy communication (Marathe et al. 2018). Importantly, autonomy must also understand when similarity is an appropriate indicator; for instance, scholars are

beginning to explore joint action dynamics beyond synchrony, such as complementarity (Dale et al. 2013) and antisynchrony (Wallot et al. 2016). Until autonomy develops these capacities to engage in timely, transparent, and socially dynamic communication—the products of team members' attempts to build common ground and achieve mutual goals—analyzing human communication in teams will enable us to push human-autonomy teaming forward.

### 3 Discussion

Human-autonomy military teams will need to leverage the kinds of flexible, adaptive interactions characteristic of human teams to maintain decisive overmatch in the dynamic future battlefield. With more effective communication assessment tools, and more datasets to draw from, a clearer picture will emerge of human-autonomy team communication, offering a better understanding of metrics or patterns that define effective teams and teaming in various contexts. The following sections will synthesize the key information presented throughout this article. First, we provide a comparison of the approaches to support the selection of the best approach (or approaches) for a given scenario. Then, we distill the directions for future research provided throughout the manuscript into four critical areas for further study.

#### 3.1 Comparing and selecting approaches

Selecting the most appropriate approaches for a given scenario will depend on characteristics of the scenario and the resources available to the researcher. To this end, Table 3 provides a comparison of the characteristics, data/resource requirements, and constraints of each approach as well as various considerations for implementing each. *Data required* describes the types of data needed to implement the approach. *Minimum sample size* provides general estimates of the minimum amount of data that should be collected and analyzed to produce usable insights using each approach. We note that those estimates are aimed at providing the reader with a general idea of the requirements of each approach; specific scenarios may warrant other requirements. Similarly, *Team size* provides the reader with estimates of the sizes of teams that are appropriate for each approach. *Required resources* outlines software or hardware packages that are associated with, or are otherwise essential to, the approach. *Curation or preprocessing required* describes, if applicable, any processing that must be done to the collected data before it can be used in analysis. *Constraints/limitations* provides insight into any constraints or limitations specific to the approach that may be relevant to its application. *Considerations for implementing approach* offers additional miscellaneous information that relates to utilizing the approach.

The relative strengths of each approach guide their selection and application for certain scenarios. For datasets that, at minimum, contain information about the senders, receivers, and timestamps of communication messages, several assessment methods are useful: aggregate communication flow, social network analysis, and relational event models. Aggregate communication flow and social network analysis can be used even without message timestamps, but due to this, they provide less nuanced data than approaches that can leverage interaction timing for more in-depth analyses. If senders, receivers, timestamps, and message content are all available, analyses of anticipatory information pushing and exploratory communication become possible, as they rely on the additional context provided by the content of the team's communication to produce findings. In addition, while message content is not strictly required for social network analysis, it improves the breadth of analyses that are possible using the approach.

It is often important to understand team dynamics over time: how teams change from one interaction to the next, how their coordination is affected by given scenarios, how they adapt their behaviors and interactions throughout the course of a task, and so on. Several approaches have a key focus on such dynamics and time-ordered interactions. Of the analyses that can leverage merely senders, receivers, and timestamps, relational event modeling is uniquely suited to provide such insights. If more data types are available such as message content, self-report data, or even interviews or system logs, the distributed dynamic approach of team cognition and the approach for quantifying exploratory communication become highly useful methods for evaluating communication interaction patterns over time. These two approaches are rooted in dynamical systems analysis, providing a theoretical foundation geared toward understanding team dynamics and cognition in depth.

For scenarios in which only one data type is available (or feasible for capture), the voice, facial expression, and linguistic synchrony approaches are especially useful. Vocal feature assessment can be performed so long as an audio stream or audio recordings are available for processing. Facial expression analysis minimally relies on video recordings, so in scenarios in which at least one camera is available to point at a crew member, this approach can be implemented, and more cameras for more crew members extends this capability. The linguistic synchrony approaches all use text data, so if a given scenario involves team interactions through a chat system, those approaches are especially useful. Spoken interactions can be transcribed to text, so even if the only available data involves audio recording, those approaches can also be implemented. It is often time-intensive to accurately transcribe audio recordings, but if it can be done, the reliance of linguistic synchrony methods on the content of the communication can result in useful insights into the team's usage of words, utterances, and syntactic structures.

**Table 3** Comparison between approaches and their characteristics

Approach: Structural analysis	Data required	Minimum sample size	Team size	Required resources	Curation or pre-processing required	Constraints/limitations	Considerations for implementing approach
Aggregate Communication Flow	Sender and receiver for each communication event	Around 15 interactions times the number of team members	Commonly 5 or fewer; larger teams are possible but resulting flow maps become more complex	See footnote <sup>1</sup> . Graphing/design software to produce flow maps (e.g., R package 'igraph' <sup>2</sup> , PowerPoint, etc.)	If recordings or observation are used, these must be coded to identify senders/receivers of each communication event	Data is aggregated, so time order of specific interactions is not considered. Thus, less nuance than more-analytical approaches	Flow maps provide quick, easily interpretable summary information about team's communication pattern
Social Network Analysis	Sender and receiver for each interaction event; timestamps and message content are optional but recommended. Alternately, network members may self-report ties in the network.	Around 15 interactions times the number of team members. Self-reports of network ties should rely on surveys from each team member.	Commonly 5 or more; larger teams enable the use of a larger number of analysis techniques	R packages such as 'igraph' <sup>2</sup> or 'sna' <sup>3,1</sup>	If recordings or observation are used, these must be coded to identify senders/receivers of each interaction event	Data is aggregated, so time order of specific interactions is not considered. Clearly delineating message recipients can be challenging in real-world team settings and can potentially impact outputs	More nuanced analysis possible compared to flow maps. May be applied to any directed interaction, including communication, resource allocation, role/task assignments, etc.
Relational Event Models	Sender and receiver for each interaction event. Timing may be exact (A-B occurred at time X, B-C occurred at time Y) or ordinal (A-B occurred first, B-C occurred second...)	Around 30 interactions, up to several thousand.	At least 3, up to several hundred	R package 'relEvent' <sup>4</sup> . Ability to identify time at which interactions occurred <sup>1</sup> .	If recordings or observation are used, these must be coded to identify senders/receivers of each interaction event	Model calculation slows with large n; some model terms assume individuals' awareness of ongoing interactions (may not be feasible for situations with many actors)	May be flexibly applied to any directed interaction, including communication, resource allocation, role/task assignments, etc.
Anticipatory Information Pushing (AIP)	Text data; self-report	Around 15 interactions times the number of team members	At least 2	Tools that may apply regular expressions, such as Excel or R, to search through interaction content are recommended. <sup>1</sup>	Interaction content may need to be translated, at least partially, and then coded to determine requests and information exchanges.	It may be complicated to fit assumptions about information requests and systematic codebooks with unstructured natural language, particularly if requests may be implied.	User should consider the contexts in which AIP is useful, when pushing is likely to occur as a result of anticipation, and when other constraints (e.g., interfaces) limit pushing or pulling possibilities.
Approach: Dynamical systems							
Distributed Dynamic Approach of Team Cognition	Variety types of communication and interaction data, including related system logs over the length of a mission plus follow-up interviews	Depends on the interaction data types and task context: can start from one team's interactions over a mission and build upon it	At least 3 members with different roles, up to several hundreds. The size depends on population of interest	Access to the communication data: Word, Excel, SPSS; Zoom recording and auto-transcriptions for interviews	Identify and clean the accessible data types and develop a codebook for the data	The output format depends on the team cognition topic and accessible relevant data.	Good for multi-stakeholder team cognition exploration. Established patterns can be used for controlled follow-up studies
Exploratory Communication Analysis	Sender, receiver, message content, and timestamps for each	Varies depending on the data type used for analysis; aggregate communication	At minimum, a dyad is required. Very large teams are	Statistical programming language such as	Requires appropriate formatting for	Requires a reasonably intact history or knowledge of prior communications	Flexible method for quantifying evolution in a team's process and

Table 3 (continued)

<p>event: any additional relevant communication dimensions with timestamps</p>	<p>requires 15 interactions per teammate at minimum and time series analysis may require many more events or more time.</p>	<p>possible although this may impact the analytical approach (e.g., structural vs. dynamical systems)</p>	<p>Matlab or R. 'Econometrics' is recommended for Matlab and 'entropy' and 'tidyverse' packages recommended for R<sup>1</sup></p>	<p>sender, receiver, timestamp, and order. Must codify exploratory communication<sup>8</sup></p>	<p>within the team. Tends to be less meaningful when the interaction data consists only of nominal tasks.</p>	<p>adaptation. Effective levels of exploration may vary based on context.</p>
<p>Approach: Emotional states Facial Expression Analysis</p>	<p>Facial electromyography, live observation, or computer-vision algorithms</p>	<p>There is no minimum. Each person's facial expression can be analyzed on a frame by frame basis but it is recommended that the sampling rate is at least 50 frames per second</p>	<p>Can be used individually or with a team of any size, as long as each relevant team member's face is recorded individually</p>	<p>No pre-processing is required. Various software tools are available that can batch-process video files on a frame by frame basis</p>	<p>Individual difference factors may predispose individuals to be more or less expressive. Therefore, baseline measures should be taken and analyses conducted on individual changes from baseline to task should be conducted.</p>	<p>This approach should be used with other measures of state or affect such as subjective or physiological measures to cross-reference results</p>
<p>Vocal Feature Assessment using Neural Networks</p>	<p>Labeled speech recordings</p>	<p>Using a trained model, less than a minute for real-time feedback</p>	<p>Can be used individually or with a team of any size, as long as member's voice is recorded individually</p>	<p>The current model parameters can be refined by additional training, however, the generic model is already trained</p>	<p>Approach is newer, so evidence of generalizability is currently limited</p>	<p>Requires additional training in novel contexts. GUI needs refining</p>
<p>Approach: Linguistic Language Style Matching</p>	<p>Speech or text data involving two or more individuals</p>	<p>No minimum size for computation. LSM is based on proportions, so researchers should determine number of utterances to produce stable proportion estimates.</p>	<p>Targeted at dyads, but method can be computed among all possible dyads, so team size not limited to 2</p>	<p>LJWC<sup>9</sup>; R package 'lingmatch'<sup>10</sup></p>	<p>Because analysis relies on transcripts, the linguistic synchrony approaches are best suited for later feedback rather than real-time insights</p>	<p>Can be used to examine synchrony for any word categories, though it is most often used to analyze function words</p>
<p>Latent Semantic Similarity</p>	<p>Speech or text data involving two or more individuals</p>	<p>Corpus needs more documents (e.g., utterances or speakers, depending on analysis) than dimensions (or "topics" that represent clusters of words that make sense together); some researchers recommend at least 300 dimensions (e.g., Dong 2005).</p>	<p>Targeted at dyads, but method can be computed among all possible dyads, so team size not limited to 2</p>	<p>R packages 'lingmatch'<sup>10</sup>, 'lsa'<sup>11</sup>, Python</p>	<p>Requires a large corpus to produce meaningful topics.</p>	<p>Highly flexible method, so results will vary depending on how semantic space is created</p>

**Table 3** (continued)

Conversation Level Syntax Similarity Metric (CASSIM)	Speech or text data involving two or more individuals	Not yet established	Targeted at dyads, but method can be computed among all possible dyads, so team size not limited to 2	Python; CASSIM <sup>12</sup>	If dialogue transcripts do not identify senders and receivers, they must be coded to do so	Very time intensive. Relatively new approach with limited evidence for utility in dialogue.	Can be used with any syntax parser, meaning it can be used to analyze any language with an existing syntax parser
--	---	---------------------	---	------------------------------	--	---	---

<sup>1</sup> This approach requires identifying senders/receivers of interactions. This can be done using a system capable of identifying senders/receivers (such as through chat logs or cueing of microphones by each member); otherwise, audio/video recording or observation can also be used to identify senders/receivers after the fact

<sup>2</sup> <https://cran.r-project.org/web/packages/igraph/index.html>

<sup>3</sup> <https://cran.r-project.org/web/packages/sna/sna.pdf>

<sup>4</sup> <https://cran.r-project.org/web/packages/relevant/index.html>

<sup>5</sup> <https://www.mathworks.com/products/econometrics.html>

<sup>6</sup> <https://cran.r-project.org/web/packages/entropy/entropy.pdf>

<sup>7</sup> <https://cran.r-project.org/web/packages/tidyverse/index.html>

<sup>8</sup> One example is to decompose the data into a state space consisting of every possible combination of senders and receivers. These may be symbolically labeled. Then, codes could be defined at certain levels of deviance (e.g., entropy) in the time series

<sup>9</sup> Pennebaker, Francis, & Booth (2001)

<sup>10</sup> <https://miserman.github.io/lingmatch/>

<sup>11</sup> <https://cran.r-project.org/web/packages/lsa/lsa.pdf>

<sup>12</sup> <https://github.com/USC-CSSL/CASSIM>

### 3.2 Critical areas for further study

The capabilities of IAs and human-autonomy teams are constantly improving. Although communication has long been a focus in the domain of human-autonomy teaming and human-robot interaction, IAs are only now becoming capable enough to understand, interact, and adapt more naturalistically with human teammates. As human-autonomy teams become better at interacting, coordinating, and achieving shared goals, it will become even more crucial to leverage communication as a window into their functioning. In many cases, the approaches described in this paper have been primarily developed and tested on human teams; consequently, the literature on human-autonomy team communication assessment, along with our understanding of best practices, will improve as the approaches are further implemented in this context.

Many of the approaches in this paper share key needs for further development spurred by current constraints and limitations. To this end, we have distilled the directions for development and future research identified throughout this paper into four critical areas for further study of communication in human-autonomy teams:

1. *More efficient data collection methods are needed.* Compared to assessment approaches that rely on physiology or behaviors, communication-based assessments can provide rich information about team interactions, but this is often at the expense of the time needed to collect and process the data. Many of the methods in this paper rely on audio transcription, task analysis, or interaction analysis, which are time-consuming and laborious. However, the development of better automated systems for transcription, event logging, and so on will speed up the rate at which communication assessments can be carried out, ideally to real time, therefore negating a primary drawback to these approaches and opening up significant avenues for understanding team interactions as they happen.
2. *How do the unique qualities of autonomy affect team interactions?* Future autonomous teammates are posited to have ever-increasing intelligence that will allow for both independent and interdependent team operations in high-risk, complex environments. It is likely that these IAs will have a variety of potential communication characteristics that deviate from standard human communication paradigms. Whatever the case, it is critical to understand how the characteristics of the autonomy can influence team interactions, to make better (and faster) predictions about the performance of human-autonomy teams.
3. *How do different communication modalities affect human-autonomy teamwork?* Human-autonomy teams can use verbal, touch/haptic, gestural, or other interactions, or even multi-modal interactions. As such, it will be important to build our understanding of how the

modalities of those interactions affect team dynamics such as common grounding, shared cognition, trust, and cohesion. This may unlock further possibilities for implementing communication assessment approaches to novel modalities and scenarios, deepening our ability to characterize how the team is performing over time.

4. *What are the patterns of communication associated with team effectiveness in different team structures and contexts?* Autonomy can play many roles within a team, and human-autonomy teams can be deployed for many scenarios. What are the most effective communication patterns for a given human-autonomy configuration? For a given scenario? Because IA capabilities are often so specific to their contexts, it is sometimes difficult to generalize findings across scenarios, but the communication assessment methods described in this article are suited to answering these questions. With more data and further implementation of these approaches, we will be better equipped to understand how communication patterns associated with a variety of team outcomes may generalize across teams and contexts.

## 4 Conclusion

In this article, we presented eleven methods for assessing team communication that are applicable to human-autonomy teaming. For each, we described the process for assessment, how the approach related to team states and outcomes, considerations for application, and current efforts to develop and apply the approaches to human-autonomy teams. Although not all methods will be useful to all human-autonomy teaming scenarios, each method presents a different window into the core functioning of a team. Many of these assessment methods have been developed primarily for human-human teams, but their applicability to human-autonomy teams is promising, especially given the goal of flexible, adaptive, human-like autonomous systems that will be integrated into future human-autonomy teams. In human-only teams, the literature relating communication, trust, and performance is abundant, whereas the literature on those relationships in human-autonomy teams has not kept pace. Therefore, our presentation of a variety of communication assessment approaches supports efforts to expand our understanding of communication, trust, team cohesion, and performance in human-autonomy teams.

The greatest advantage of analyzing communication is that it can be measured unobtrusively, relying on recordings of the team's speech, chat messages, vocal features, or other data that can be captured real time. This is key to addressing the latest needs for naturalistic, objective, and continuous assessments for human-autonomy teams. Future research into



communication in military human-autonomy teams should leverage the assessment methods discussed in this article, yielding valuable insights into the four critical areas for future research and paving the way for more capable, flexible, and effective teams.

**Acknowledgements** The authors would like to thank Dr. Jason S. Metcalfe for feedback on revisions to the manuscript.

**Funding** This research was supported in part by an appointment to the US Army Research Postdoctoral Fellowship Program administered by the Oak Ridge Associated Universities through a cooperative agreement with the US Army Combat Capabilities Development Command (CCDC) Army Research Laboratory (ARL). Research was sponsored by CCDC ARL and was accomplished under Cooperative Agreement No. W911NF-18-2-0230. Arizona State University conducted research under the partial support of Army Research Laboratory Cooperative Agreement No. W911NF-18-2-0271.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Disclaimer** The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the DEVCOM Army Research Laboratory or the US Government.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abraham R, Shaw CD (1992) *Dynamics: the geometry of behavior*, 2nd edn. Addison-Wesley, Boston
- Ahronson A, Cameron JE (2007) The nature and consequences of group cohesion in a military sample. *Mil Psychol* 19(1):9–25
- Altmann GT, Kamide Y (2007) The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *J Mem Lang* 57(4): 502–518
- Amazeen PG (2018) From physics to social interactions: scientific unification via dynamics. *Cogn Syst Res* 52:640–657. <https://doi.org/10.1016/j.cogsys.2018.07.033>
- Babcock MJ, Ta VP, Ickes W (2014) Latent semantic similarity and language style matching in initial dyadic interactions. *J Lang Soc Psychol* 33(1):78–88
- Baker AL, Phillips EK, Ullman D, Keebler JR (2018) Toward an understanding of trust repair in human-robot interaction: current research and future directions. *ACM Trans Interactive Intell Syst* 8(4):1–30. <https://doi.org/10.1145/3181671>
- Baker AL, Schaefer KE, Hill SG (2019) Teamwork and communication methods and metrics for human–autonomy teaming. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); Report No.: ARL-TR-8844.
- Baker AL, Brewer RW, Schaefer KE (2020a) Development and usability assessment of the Realtime Event, Flow, and Coordination Tool (REFLECT). Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); Report No.: ARL-TR-9012.
- Baker AL, Fitzhugh SM, Forster DE, Krausman A, Scharine, A, Schaefer KE (2020b) Team trust in human-autonomy teams: analysis of crew communication during manned-unmanned gunnery operations. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); Report No.: ARL-TR-8969.
- Balkundi P, Harrison DA (2006) Ties, leaders, and time in teams: Strong inference about network structure's effects on team viability and performance. *Acad Manag J* 49(1):49–68
- Baltrušaitis T, Zadeh A, Lim YC, Morency L-P (2018) OpenFace 2.0: facial behavior analysis toolkit. *IEEE International Conference on Automatic Face and Gesture Recognition*
- Banerjee D, Islam K, Mei G, Xiao L, Zhang G, Xu R et al (2017) A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. *Proc IEEE Int Conf Data Min.* <https://doi.org/10.1109/ICDM.2017.10>
- Barber DJ, Abich J, Phillips E, Talone AB, Jentsch F, Hill SG (2015) Field assessment of multimodal communication for dismounted human-robot teams. *Proc Human Factors Ergon Soc Ann Meet* 59(1):921–925. <https://doi.org/10.1177/1541931215591280>
- Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. *Proc Natl Acad Sci* 101(11):3747–3752
- Barth S, Schraagen JM, Schmettow M (2015) Network measures for characterising team adaptation processes. *Ergonomics* 58(8):1287–1302
- Bartlett MS, Littlewort G, Fasel I, Chenu J, Kanda T, Ishiguro H, Movellan JR (2004) Towards social robots: automatic evaluation of human-robot interaction by face detection and expression classification. In *Proc IEEE Int Conf Systems, Man, and Cybernetics* p. 592–597
- Batrinca L, Stratou G, Shapiro A, Morency LP, Scherer S (2013) Cicero-towards a multimodal virtual audience platform for public speaking training. In *International Workshop on Intelligent Virtual Agents* (p 116–128). Springer
- Beal DJ, Cohen RR, Burke MJ, McLendon CL (2003) Cohesion and performance in groups: a meta-analytic clarification of construct relations. *J Appl Psychol* 88(6):989–1004
- Beilock SL, Goldin-Meadow S (2010) Gesture changes thought by grounding it in action. *Psychol Sci* 21(11):1605–1610
- Berlo D (1960) *Process of communication: an introduction to theory and practice*. Harcourt School, San Diego
- Bisk, Y., Yuret, D., & Marcu, D. (2016). *Natural language communication with robots*. Paper presented at the Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, California.
- Bock JK (1982) Toward a cognitive psychology of syntax: information processing contributions to sentence formulation. *Psychol Rev* 89(1):1–47
- Boghtrati R, Dehghani M (2018) Follow my language! Effect of power relations on syntactic alignment. Paper presented at the CogSci.
- Boghtrati R, Hoover J, Johnson KM, Garten J, Dehghani M (2018) Conversation level syntax similarity metric. *Behav Res Methods* 50(3):1055–1073

- Borgatti SP, Everett MG, Freeman LC (2002) Ucinet for windows: software for social network analysis. Analytic Technologies, Harvard, p 2006
- Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. *Science* 323(5916):892–895
- Borgatti SP, Everett MG, Johnson JC (2018) Analyzing social networks. Sage, Thousand Oaks
- Bou-Ghazale SE, Hansen JHL (2000) A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Trans Speech Audio Process* 8:429–442. <https://doi.org/10.1109/89.848224>
- Brandes U, Lerner J, Snijders TA (2009) Networks evolving step by step: statistical analysis of dyadic event data. In: 2009 international conference on advances in social network analysis and mining. IEEE, pp 200–205
- Brewer RW, Cerame E, Pursel ER, Zimmermann A, Schaefer KE (2018) Manned-unmanned teaming: US army robotic wingman vehicles. In: In: International Conference on Applied Human Factors and Ergonomics, 2018. Advances in Intelligent Systems and Computing. Springer, Cham, pp 89–100. [https://doi.org/10.1007/978-3-319-94223-0\\_9](https://doi.org/10.1007/978-3-319-94223-0_9)
- Brewer RW, Walker AJ, Pursel ER, Cerame EJ, Baker AL, Schaefer KE (2019) Assessment of manned-unmanned team performance: comprehensive after-action review technology development. In: International Conference on Applied Human Factors and Ergonomics, Washington, DC, 2019. Advances in human factors in robots and unmanned systems. Springer International Publishing, New York City, pp 119–130
- Brown TM, Miller CE (2000) Communication networks in task-performing groups: effects of task complexity, time pressure, and interpersonal dominance. *Small Group Res* 31(2):131–157
- Burke C, Sims DE, Lazzara EH, Salas E (2007) Trust in leadership: a multi-level review and integration. *Leadersh Q* 18:606–632. <https://doi.org/10.1016/j.leaqua.2007.09.006>
- Burkhardt F, Paeschke A, Rolfes M, Sendlmeier WF, Weiss B (2005) A database of German emotional speech. *Interspeech* 33:10, 12. <https://doi.org/10.1109/35.339860>
- Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS (2008) IEMOCAP: interactive emotional dyadic motion capture database. *Lang Resour Eval* 42(4):335–359. <https://doi.org/10.1007/s10579-008-9076-6>
- Butts CT (2008) A relational event framework for social action. *Sociol Methodol* 38(1):155–200
- Cannizzaro M, Harel B, Reilly N, Chappell P, Snyder PJ (2004) Voice acoustical measurement of the severity of major depression. *Brain Cogn* 56:30–35. <https://doi.org/10.1016/j.bandc.2004.05.003>
- Casale, S., Russo, A., Scebbra, G., & Serrano, S. (2008). Speech emotion classification using Machine Learning algorithms. *Proceedings - IEEE International Conference on Semantic Computing 2008, ICSC 2008*. <https://doi.org/10.1109/ICSC.2008.43>
- Chen JYC, Lakhmani SG, Stowers K, Selkowitz AR, Wright JL, Barnes M (2018) Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theor Issues Ergon Sci* 19:259–282. <https://doi.org/10.1080/1463922X.2017.1315750>
- Chen S, Andrienko N, Andrienko G, Adilova L, Barlet J, Kindermann J et al (2019) LDA ensembles for interactive exploration and categorization of behaviors. *IEEE Trans Vis Comput Graph*
- Chiocchio F, Essiembre H (2009) Cohesion and performance: a meta-analytic review of disparities between project teams, production teams, and service teams. *Small Group Res* 40(4):382–420
- Chollet, M., Wörtwein, T., Morency, L. P., Shapiro, A., & Scherer, S. 2015. Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing(pp. 1143-1154).
- Clark HH, Brennan SE (1991) Grounding in communication. In: Perspectives on socially shared cognition. American Psychological Association, Washington, DC, pp 127–149
- Colquitt JA, Scott BA, LePine JA (2007) Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *J Appl Psychol* 92(4):909–927. <https://doi.org/10.1037/0021-9010.92.4.909>
- Cook Z, Franks DW, Robinson EJ (2013) Exploration versus exploitation in polydomous ant colonies. *J Theor Biol* 323:49–56
- Cooke NJ (2015) Team cognition as interaction. *Curr Dir Psychol Sci* 24(6):415–419
- Cooke NJ, Gorman JC (2009) Interaction-based measures of cognitive systems. *J Cogn Eng Decis Making* 3:27–46
- Cooke, N. J., & Shope, S. M. (2004). Designing a synthetic task environment. *Scaled worlds: Development, validation, and application*, 263-278.
- Cooke NJ, Gorman JC, Myers C, Duran J (2012) Theoretical underpinning of interactive team cognition
- Cooke NJ, Gorman JC, Myers CW, Duran JL (2013) Interactive team cognition. *Cogn Sci* 37(2):255–285. <https://doi.org/10.1111/cogs.12009>
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Int J Complex Syst* 1695(5):1–9
- Dale, R., Fusaroli, R., Døjbak Håkonsson, D. D., Healey, P., Mønster, D., McGraw, J., ... & Tylén, K. (2013). Beyond synchrony: complementarity and asynchrony in joint action. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).
- Dapretto M, Bookheimer SY (1999) Form and content: dissociating syntax and semantics in sentence comprehension. *Neuron* 24(2):427–432
- De Jong BA, Elfring T (2010) How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Acad Manag J* 53(3):535–549. <https://doi.org/10.5465/amj.2010.51468649>
- de Visser EJ, Peeters MM, Jung MF, Kohn S, Shaw TH, Pak R, Neerincx MA (2019) Towards a theory of longitudinal trust calibration in human-robot teams. *Int J Soc Robot* 12:1–20
- Demir M, McNeese NJ, Cooke NJ (2017) Team situation awareness within the context of human-autonomy teams. *J Cogn Syst Res* 46(1):3–12
- Demir M, Likens AD, Cooke NJ, Amazeen PG, McNeese NJ (2018) Team coordination and effectiveness in human-autonomy teaming. *IEEE Trans Human Mach Syst* 49(2):150–159
- Demir M, McNeese NJ, She M, Cooke NJ (2019) Team coordination of team situation awareness in human-autonomy teaming. *Proc Human Factors Ergon Soc Ann Meet* 63(1):146–147
- Demir M, McNeese NJ, Cooke NJ (2020) Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *Int J Hum Comput Stud* 140:102436. <https://doi.org/10.1016/j.ijhcs.2020.102436>
- D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, K. Georgila, J. Gratch, A. Hartholt, M. Lhommel, et al. Sim-sensei kiosk: a virtual human interviewer for healthcare decision support. In Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems, pages 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems, 2014
- Dong A (2005) The latent semantic approach to studying design team communication. *Des Stud* 26(5):445–461
- Dong A, Hill AW, Agogino AM (2004) A document analysis method for characterizing design team performance. *J Mech Des* 126(3):378–385
- Dong Y, Hu Z, Uchimura K, Murayama N (2011) Driver inattention monitoring system for intelligent vehicles: a review. *IEEE Trans Intell Transp Syst* 12:596–614
- Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP (2003) The role of trust in automation reliance. *Int J Hum-Comput Stud* 58(6):697–718. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)

- Ekman P, Friesen W (1978) Facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists Press, Palo Alto
- El Ayadi M, Kamel MS, Karray F (2011) Survey on speech emotion recognition: features, classification schemes, and databases. *Pattern Recogn* 44:572–587. <https://doi.org/10.1016/j.patcog.2010.09.020>
- Elliott LR, Hill SG, Barnes M. (2016). Gesture-based controls for robots: overview and implications for use by soldiers. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); Report No.: ARL-TR-7715.
- Endsley MR (2015) Situation awareness misconceptions and misunderstandings. *J Cogn Eng Decis Making* 9:4–32. <https://doi.org/10.1177/1555343415572631>
- Endsley MR (2017) From here to autonomy: lessons learned from human–automation research. *Hum Factors* 59(1):5–27. <https://doi.org/10.1177/0018720816681350>
- Entin EE, Serfaty D (1999) Adaptive team coordination. *Hum Factors* 41(2):312–325
- Evangelopoulos N, Zhang X, Prybutok VR (2012) Latent semantic analysis: five methodological recommendations. *Eur J Inf Syst* 21(1):70–86
- Finan PH, Hessler EE, Amazeen PG, Butner J, Zautra AJ, Tennen H (2010) Oscillations in daily pain prediction accuracy. *Nonlinear Dyn Psychol Life Sci* 14(1):27–46
- Fischer U, McDonnell L, Orasanu J (2007) Linguistic correlates of team performance: toward a tool for monitoring team functioning during space missions. *Aviat Space Environ Med* 78:B86–B95
- Fiscus JG, Garofolo JS, Lamel LF, Fischer WM, Fiscus JG, Pallett DS, Dahlgren NL (1993) DARPA TIMIT acoustic-phonetic continuous speech corpus CD-ROM. NASA STI/Recon Technical Report N 27: 1457–1470. <https://doi.org/10.1081/PLN-200025865>
- Foltz, P. W., Laham, D., & Derr, M. (2003). *Automated speech recognition for modeling team performance*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.
- Foltz, P. W., Martin, M. J., Abdelali, A., Rosenstein, M., & Oberbreckling, R. (2006). *Automated team discourse modeling: test of performance and generalization*. Paper presented at the Proceedings of the 28th Annual Cognitive Science Conference.
- Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Networks* 1(3):215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Friedberg, H., Litman, D., & Paletz, S. B. (2012). *Lexical entrainment and success in student engineering groups*. Paper presented at the 2012 IEEE Spoken Language Technology Workshop (SLT).
- Funke GJ, Knott BA, Salas E, Pavlas D, Strang AJ (2012) Conceptualization and measurement of team workload: a critical need. *Hum Factors* 54(1):36–51
- Fusaroli R, Bahrami B, Olsen K, Roepstorff A, Rees G, Frith C, Tylén K (2012) Coming to terms: quantifying the benefits of linguistic coordination. *Psychol Sci* 23(8):931–939
- Galati A, Brennan SE (2014) Speakers adapt gestures to addressees' knowledge: implications for models of co-speech gesture. *Lang Cogn Neurosci* 29(4):435–451
- Geisler C, Swarts J (2019) *Coding streams of language: techniques for the systematic coding of text, talk, and other verbal data*. WAC Clearinghouse
- Gonzales AL, Hancock JT, Pennebaker JW (2010) Language style matching as a predictor of social dynamics in small groups. *Commun Res* 37(1):3–19
- Goodwin C (1986) Gestures as a resource for the organization of mutual orientation. *Semiotica* 62(1-2):29–50
- Gorman JC, Cooke NJ (2011) Changes in team cognition after a retention interval: the benefits of mixing it up. *J Exp Psychol Appl* 17(4):303
- Gorman, J. C., Foltz, P. W., Kiekel, P. A., Martin, M. J., & Cooke, N. J. (2003). *Evaluation of latent semantic analysis-based measures of team communications content*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society annual meeting.
- Gorman JC, Cooke NJ, Amazeen PG (2010) Training adaptive teams. *Hum Factors* 52(2):295–307. <https://doi.org/10.1177/0018720810371689>
- Gorman JC, Dunbar TA, Grimm D, Gipson CL (2017) Understanding and modeling teams as dynamical systems. *Front Psychol* 8:1053. <https://doi.org/10.3389/fpsyg.2017.01053>
- Gorman JC, Demir M, Cooke NJ, Grimm DA (2019) Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: a layered dynamics approach. *Ergonomics* 62(5):629–643
- Grossman R, Feitosa J (2018) Team trust over time: modeling reciprocal and contextual influences in action teams. *Hum Resour Manag Rev* 28(4):395–410
- Gu, H., & Ji, Q. (2004). An automated face reader for fatigue detection. In: FGR, pp. 111–116.
- Hagoort P (2003) Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J Cogn Neurosci* 15(6):883–899
- Hancock PA (2017) Imposing limits on autonomous systems. *Ergonomics* 60:284–291. <https://doi.org/10.1080/00140139.2016.1190035>
- Hancock PA, Billings DR, Schaefer KE, Chen JYC, de Visser EJ, Parasuraman R (2011) A meta-analysis of factors affecting trust in human-robot interaction. *Hum Factors* 53(5):517–527. <https://doi.org/10.1177/0018720811417254>
- Handcock, M. S., Hunter D. R., Butts C. T., Goodreau S. M., Krivitsky P. N., Bender-de Moll S., & Morris, M. (2008). statnet: software tools for the statistical analysis of network data. The Statnet Project (<http://www.statnet.org>), <https://CRAN.R-project.org/package=statnet>. Hansen,
- J. H. L. Hansen, & Patil, S. (2007). Speech under stress: Analysis, modeling and recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [https://doi.org/10.1007/978-3-540-74200-5\\_6](https://doi.org/10.1007/978-3-540-74200-5_6)
- Harris, J., & Barber, D. Speech and gesture interfaces for squad-level human-robot teaming. In: Unmanned Systems Technology XVI, 2014/06/03/ 2014. International Society for Optics and Photonics, p 90840B. doi:<https://doi.org/10.1117/12.2052961>
- Heuer K, Müller-Frommeyer LC, Kauffeld S (2020) Language matters: the double-edged role of linguistic style matching in work groups. *Small Group Res* 51(2):208–228
- Hill, S. G. (2017). Human-robot bi-directional communications literature overview (DST-Group-TR-3416).
- Hills TT, Todd PM, Lazer D, Redish AD, Couzin ID, Cognitive Search Research Group (2015) Exploration versus exploitation in space, mind, and society. *Trends Cogn Sci* 19(1):46–54
- Ho, N., Johnson, W., Panesar, K., Wakeland, K., Sadler, G., Wilson, N., Nguyen, B., Lachter, J., & Brandt, S. (2017). Application of human-autonomy teaming to an advanced ground station for reduced crew operations. *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, 1–4.
- Hoff K, Bashir M (2015) Trust in automation: integrating empirical evidence on factors that influence trust. *Hum Factors* 57(3):407–434
- Huang, C. W., & Narayanan, S. S. (2017). Deep convolutional recurrent neural network with attention mechanism for robust speech emotion recognition. *Proceedings - IEEE International Conference on Multimedia and Expo*, 583–588. <https://doi.org/10.1109/ICME.2017.8019296>
- Huang Y, Xiao J, Tian K, Wu A, Zhang G (2019) Research on robustness of emotion recognition under environmental noise conditions. *IEEE Access* 7:142009–142021. <https://doi.org/10.1109/access.2019.2944386>
- Huang, L., Cummings, M., & Ono, M. (2020a). A mixed analysis of influencing factors for trust in a risk-aware autonomy experiment.

*Proceedings of Human Factors and Ergonomics Society Annual Conference.*

- Huang L, Cooke NJ, Gutzwiller RS, Berman S, Chiou EK, Demir M, Zhang W (2020b) Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In: Nam CS, Lyons JB (eds) *Trust in Human-Robot Interaction*. Academic Press, Cambridge, pp 301–319
- Huang, L., Cooke, N., Johnson, C., Lematta, G., Bhatti, S., Barnes, M., & Holder, E. (2020c). *Human-autonomy teaming: interaction metrics and models for Next Generation Combat Vehicles*. Arizona State University and Army Research Laboratory First Year Report.
- Hung H, Gatica-Perez D (2010) Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Trans Multimed* 12: 563–575. <https://doi.org/10.1109/TMM.2010.2055233>
- Johnson M, Vera A (2019) No AI is an island: the case for teaming intelligence. *AI Mag* 40:16. <https://doi.org/10.1609/aimag.v40i1.2842>
- Johnson B, Floyd MW, Coman A, Wilson MA, Aha DW (2018) Goal reasoning and trusted autonomy. In: Abbass H, Scholtz J, Reid DJ (eds) *Foundations of Trusted Autonomy*. Springer, Cham. Coding streams of language: techniques for the systematic
- Johnson, C., Lematta, G., Huang, L., Holder, E., Bhatti, S., & Cooke, N. (2020a). An interaction taxonomy of human-agent teaming in Next Generation Combat Vehicle Systems. *Proceedings of the AHFE 2020 International Conference on Human Factors in Robots and Unmanned Systems, 1210*, 10–16.
- Johnson, C. J., Demir, M., Zabala, G. M., He, H., Grimm, D. A., Radigan, C., Wolff, A. T., Cooke, N. J., McNeese, N. J., & Gorman, J. C. (2020b). Training and verbal communications in human-autonomy teaming under degraded conditions. In *2020 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* (pp. 53–58). IEEE.
- Joshi A, Lazarova MB, Liao H (2008) Getting everyone on board: the role of inspirational leadership in geographically dispersed teams. *Organ Sci* 20(1):240–252. <https://doi.org/10.1287/orsc.1080.0383>
- Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. *J Artif Intell Res* 4:237–285
- Kattoju, R. K., Barber, D. J., Abich, J., & Harris, J. (2016). Technological evaluation of gesture and speech interfaces for enabling dismounted soldier-robot dialogue. In: *Unmanned Systems Technology XVIII, 2016/05/13/ 2016*. International Society for Optics and Photonics, p 98370N. doi:<https://doi.org/10.1117/12.2223894>
- Katz N, Lazer D, Arrow H, Contractor N (2004) Network theory and small groups. *Small Group Res* 35(3):307–332. <https://doi.org/10.1177/1046496404264941>
- Kiekel PA, Cooke NJ, Foltz PW, Shope SM (2001) Automating measurement of team cognition through analysis of communication data. In: Smith MJ, Koubek RJ, Salvendy G, Harris D (eds) *Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents, and Virtual Reality*. Erlbaum Associates, Mahwah, pp 1382–1386
- Klein G, Feltoovich PJ, Bradshaw JM, Woods DD (2005) Common ground and coordination in joint activity. *Organ Simul* 53:139–184
- Knoeferle P, Kreysa H (2012) Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Front Psychol* 3:538
- Koolagudi SG, Rao KS (2012) Emotion recognition from speech: a review. *Int J Speech Technol* 15:99–117. <https://doi.org/10.1007/s10772-011-9125-1>
- Krackhardt, D. (2014). “A preliminary look at accuracy in egonets” In *Contemporary Perspectives on Organizational Social Networks*. Emerald Group Publishing Limited.
- Krausman A, Kelley T, McGhee S, Schaefer KE, Fitzhugh S (2019) Using Dragon for speech-to-text transcription in support of human-autonomy teaming research. U.S. Army Combat Capabilities Development Command, Army Research Laboratory, Aberdeen Proving Ground, MD, United States (ARL-TN-0978)
- Lakhmani S., Abich J., Barber D., Chen J. (2016) A proposed approach for determining the influence of multimodal robot-of-human transparency information on human-agent teams. In: Schmorrow D., Fidopiastis C. (eds) *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience*. AC 2016. Lecture Notes in Computer Science, vol 9744. Springer, Cham
- Landauer TK, Dumais ST (1997) A solution to Plato’s problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol Rev* 104(2):211–240
- Landauer TK, Foltz PW, Laham D (1998) An introduction to latent semantic analysis. *Discourse Process* 25(2-3):259–284
- Landauer TK, McNamara DS, Dennis S, Kintsch W (2013) *Handbook of latent semantic analysis*. Psychology Press, Hove
- Lee JD, Kolodge K (2018) Understanding attitudes towards self-driving vehicles: quantitative analysis of qualitative data. *Proc Human Factors Ergon Soc Ann Meet* 62:1399–1403
- Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. *Hum Factors* 46(1):50–80. [https://doi.org/10.1518/hfes.46.1.50\\_30392](https://doi.org/10.1518/hfes.46.1.50_30392)
- Leenders RTA, Contractor NS, DeChurch LA (2016) Once upon a time: understanding team processes as relational event networks. *Organ Psychol Rev* 6(1):92–115
- Lewis M, Sycara K, Walker P (2018) The role of trust in human-robot interaction. In: Abbass HA, Scholz J, Reid DJ (eds) *Foundations of Trusted Autonomy*. Springer International Publishing, Cham, pp 135–159. [https://doi.org/10.1007/978-3-319-64816-3\\_8](https://doi.org/10.1007/978-3-319-64816-3_8)
- Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, Bohnen J, Orser B, Doran D, Grober E (2004) Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Safety Health Care* 13(5):330–334. <https://doi.org/10.1136/qshc.2003.008425>
- Litman, D., Paletz, S., Rahimi, Z., Allegretti, S., & Rice, C. (2016). The teams corpus and entrainment in multi-party spoken dialogues. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1421–1431).
- Liu ZT, Wu M, Cao WH, Mao JW, Xu JP, Tan GZ (2018) Speech emotion recognition based on feature selection and extreme learning machine decision tree. *Neurocomputing* 273:271–280. <https://doi.org/10.1016/j.neucom.2017.07.050>
- MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: the hidden cost of team cognition.
- Marathe AR, Schaefer KE, Evans AW, Metcalfe JS (2018) Bidirectional communication for effective human-agent teaming. In: Chen J, Fragomeni G (eds) *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation*. VAMR 2018. Lecture Notes in Computer Science, vol 10909. Springer, Cham
- Marge M, Bonial C, Pollard KA, Artstein R, Byrne B, Hill SG et al (2016) Assessing agreement in human-robot dialogue strategies: a tale of two wizards *Intelligent Virtual Agents*. Springer, Cham, pp 484–488
- Marks MA, Mathieu JE, Zaccaro SJ (2001) A temporally based framework and taxonomy of team processes. *Acad Manag Rev* 26(3): 356–376. <https://doi.org/10.2307/259182>
- Martin, M., & Foltz, P. (2004). *Automated team discourse annotation and performance prediction using LSA*. Paper presented at the Proceedings of HLT-NAACL 2004: Short Papers.
- McAllister D (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad Manag J* 38:24–59. <https://doi.org/10.2307/256727>
- McNeese NJ, Demir M, Cooke NJ, Myers C (2018) Teaming with a synthetic teammate: insights into human-autonomy teaming. *Hum Factors* 60:262–273. <https://doi.org/10.1177/0018720817743223>
- McNeese, N., Demir, M., Chiou, E., Cooke, N., & Yanikian, G. (2019). *Understanding the role of trust in human-autonomy teaming*. Paper presented at the Proceedings of the 52nd Hawaii International Conference on System Sciences, Grand Wailea, Maui.

- Mercado JE, Rupp ME, Chen JYC, Barnes MJ, Barber D, Procci K (2016) Intelligent agent transparency in human-agent teaming for multi-UxV management. *Hum Factors* 58(3):401–415
- Mesmer-Magnus JR, Dechurch LA (2009) Information sharing and team performance: a meta-analysis. *J Appl Psychol* 94:535–546. <https://doi.org/10.1037/a0013773>
- Mesmer-Magnus JR, DeChurch LA, Jimenez-Rodriguez M, Wildman JL, Shuffler M (2011) A meta-analytic investigation of virtuality and information sharing in teams. *Organ Behav Hum Decis Process* 115(2): 214–225. <https://doi.org/10.1016/j.obhdp.2011.03.002>
- Moreno JL, Jennings HH (1938) Statistics of social configurations. *Sociometry* 1:342–374
- Munson, S. A., Kervin, K., & Robert Jr, L. P. (2014). *Monitoring email to indicate project team performance and mutual attraction*. Paper presented at the Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing.
- Nenkova, A., Gravano, A., & Hirschberg, J. (2008). *High frequency word entrainment in spoken dialogue*. Paper presented at the Proceedings of ACL-08: HLT, Short Papers.
- Neubauer, C., Chollet, M., Mozgai, S., Dennison, M., Khooshabeh, P., & Scherer, S. 2017. The relationship between task-induced stress, vocal changes, and physiological state during a dyadic team task. In *Proc. of the 19th ACM Int. Conf. on Multimodal Interaction*. ACM, 426–432
- Niebles JC, Wang H, Fei-Fei L (2008) Unsupervised learning of human action categories using spatial-temporal words. *Int J Comput Vis* 79(3):299–318
- Niederhoffer KG, Pennebaker JW (2002) Linguistic style matching in social interaction. *J Lang Soc Psychol* 21(4):337–360
- Odean R, Nazareth A, Pruden SM (2015) Novel methodology to examine cognitive and experiential factors in language development: combining eye-tracking and LENA technology. *Front Psychol* 6:1266
- Oliver LW, Harman J, Hoover E, Hayes SM, Pandhi NA (1999) A quantitative integration of the military cohesion literature. *Mil Psychol* 11(1):57–83
- Ososky S, Schuster D, Jentsch F, Fiore S, Shumaker R, Lebiere C et al (2012) The importance of shared mental models and shared situation awareness for transforming robots from tools to teammates. In: *Unmanned systems technology XIV*, vol 8387. International Society for Optics and Photonics, p 838710
- Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. *IEEE Trans Syst Man Cybern A: Syst Humans* 30:286–297. <https://doi.org/10.1109/3468.844354>
- Parra F, Miljkovitch R, Persiaux G, Morales M, Scherer S (2017) The multimodal assessment of adult attachment security: developing the biometric attachment test. *J Med Internet Res* 19(4):e100
- Pennebaker JW, Francis ME, Booth RJ (2001) *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates 71(2001):2001
- Phillips E, Ososky S, Grove J, Jentsch F (2011) From tools to teammates: toward the development of appropriate mental models for intelligent robots. *Proc Human Factors Ergon Soc Ann Meet* 55:1491–1495. <https://doi.org/10.1177/1071181311551310>
- Pilny A, Yahja A, Poole MS, Dobosh M (2014) A dynamic social network experiment with multi-team systems. In: 2014 IEEE fourth international conference on big data and cloud computing. IEEE, pp 587–593
- Pilny A, Schecter A, Poole MS, Contractor N (2016) An illustration of the relational event model to analyze group interaction processes. *Group Dyn Theory Res Pract* 20(3):181
- Pilny A, Proulx JD, Dinh L, Bryan AL (2017) An adapted structural framework for the emergence of communication networks. *Commun Stud* 68(1):72–94
- Plant KL, Stanton NA (2016) Distributed cognition in search and rescue: loosely coupled tasks and tightly coupled roles. *Ergonomics* 59(10): 1353–1376
- Quintane E, Conaldi G, Tonellato M, Lomi A (2014) Modeling relational events: a case study on an open source software project. *Organ Res Methods* 17(1):23–50
- Rahimi, Z., Kumar, A., Litman, D. J., Paletz, S., & Yu, M. (2017). *Entrainment in multi-party spoken dialogues at multiple linguistic levels*. Paper presented at the INTERSPEECH.
- Richardson BH, McCulloch KC, Taylor PJ, Wall HJ (2019) The cooperation link: power and context moderate verbal mimicry. *J Exp Psychol Appl* 25(1):62–76
- Roberts APJ, Stanton NA, Fay DT, Pope KA (2019) The effects of team co-location and reduced crewing on team communication characteristics. *Appl Ergon* 81:102875. <https://doi.org/10.1016/j.apergo.2019.102875>
- Rolf M, Steil JJ, Gienger M (2011) Online goal babbling for rapid bootstrapping of inverse models in high dimensions. In: 2011 IEEE international conference on development and learning (ICDL), vol 2. IEEE, pp 1–8
- Sacks H, Schegloff EA, Jefferson G (1974) A simplest systematics for the organization of turn-taking for conversation language. 50:696–735. <https://doi.org/10.2307/412243>
- Salas E, Cannon-Bowers JA (2000) The anatomy of team training. *Train Retrain*:312–335
- Salas E, Sims DE, Burke CS (2005) Is there a “Big Five” in Teamwork? *Small Group Res* 36:555–599. <https://doi.org/10.1177/1046496405277134>
- Salas E, Wilson KA, Murphy CE, King H, Salisbury M (2008) Communicating, coordinating, and cooperating when lives depend on it: tips for teamwork. *Jt Comm J Qual Patient Saf* 34:333–341. [https://doi.org/10.1016/S1553-7250\(08\)34042-2](https://doi.org/10.1016/S1553-7250(08)34042-2)
- Salas E, Grossman R, Hughes AM, Coultas CW (2015) Measuring team cohesion: observations from the science. *Hum Factors* 57:365–374. <https://doi.org/10.1177/0018720815578267>
- Schaefer, K. E., & Straub, E. R. (2016). *Will passengers trust driverless vehicles? Removing the steering wheel and pedals*. Paper presented at the 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA).
- Schaefer KE, Chen JYC, Szalma JL, Hancock PA (2016) A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. *Hum Factors* 58(3):377–400. <https://doi.org/10.1177/0018720816634228>
- Schaefer KE, Straub ER, Chen JYC, Putney J, Evans AW (2017) Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cogn Syst Res* 46:26–39. <https://doi.org/10.1016/j.cogsys.2017.02.002>
- Schaefer, K. E., Baker, A. L., Brewer, R. W., Patton, D., Canady, J., & Metcalfe, J. S. Assessing multi-agent human-autonomy teams: US Army Robotic Wingman gunnery operations. In: George T, Islam MS (eds) *SPIE Defense + Commercial Sensing*, Baltimore, MD, 2019a. International Society for Optics and Photonics, p 109822B. doi:<https://doi.org/10.1117/12.2519302>
- Schaefer, K. E., Brewer, R. W., Baker, A. L., Pursel, E. R., Gipson, B., Ratka, S., Giacchi, J., Cerame, E., & Pirozzo, K.. US Army Wingman Joint Capability Technology Demonstration (JCTD): initial soldier and marine feedback on manned-unmanned gunnery operations. U.S. Army Combat Capabilities Development Command Army Research Laboratory. Aberdeen Proving Ground, United States, 2019b.
- Schaefer KE, Perelman BS, Gremillion GM, Marathe AR, Metcalfe JS (2020) A roadmap for developing team trust metrics for human-autonomy teams. In: Nam CS, Lyons J (eds) *Trust in Human-Robot Teaming*. Elsevier, Amsterdam (Netherlands)
- Schecter, A. M. (2017). *It's about time: theorizing the antecedents and outcomes of dynamic processes in teams and multi-team systems* (Doctoral dissertation, Northwestern University).

- Schechter, A., & Contractor, N. (2016). A multi relational event network approach to studying channel selection dynamics in teams. In *Academy of Management Proceedings* (Vol. 2016, No. 1, p. 16663). Briarcliff Manor, NY 10510: Academy of Management.
- Scherer S, Lucas GM, Gratch J, Rizzo AS, Morency LP (2016) Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews. *IEEE Trans Affect Comput* 7(1):59–73
- Scheutz, M., Dame, N., Schermerhorn, P., & Kramer, J. (2006). The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of ACM SIGCHI/SIGART Conf. Human-Robot Interaction*, pp. 226–233.
- Schuller, B., Rigoll, G., & Lang, M. (2003). Hidden Markov model-based speech emotion recognition. *Proceedings - IEEE International Conference on Multimedia and Expo* <https://doi.org/10.1109/ICME.2003.1220939>
- Scissors, L. E., Gill, A. J., Geraghty, K., & Gergle, D. (2009). In *CMC we trust: the role of similarity*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
- Semin, G. R. (2007). Grounding communication: synchrony.
- Shannon CE (1948) A Mathematical theory of communication. *Bell Syst Tech J* 27:379–423. <https://doi.org/10.1145/584091.584093>
- Shaw ME (1964) Communication Networks. In: Berkowitz L (ed) *Advances in Experimental Social Psychology* (Vol. 1, pp. 111–147). Academic Press, Cambridge. [https://doi.org/10.1016/S0065-2601\(08\)60050-7](https://doi.org/10.1016/S0065-2601(08)60050-7)
- Sheridan TB, Verplank W (1978) Human and computer control of undersea teleoperators. Man-Machine Systems Laboratory, Department of Mechanical Engineering, MIT, Cambridge
- Sparowe RT, Liden RC, Wayne SJ, Kraimer ML (2001) Social networks and the performance of individuals and groups. *Acad Manag J* 44(2):316–325
- Stanton NA, Walker GH, Sorensen LJ (2012) It's a small world after all: contrasting hierarchical and edge networks in a simulated intelligence analysis task. *Ergonomics* 55(3):265–281
- Stanton NA, Stewart R, Harris D, Houghton RJ, Baber C, McMaster R et al (2006) Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics* 49(12–13):1288–1311
- Staudte M, Crocker MW, Heloir A, Kipp M (2014) The influence of speaker gaze on listener comprehension: Contrasting visual versus intentional accounts. *Cognition* 133(1):317–328
- Stowers K, Kasdaglis N, Newton O, Lakhmani S, Wohleber R, Chen J (2016) Intelligent agent transparency: the design and evaluation of an interface to facilitate human and intelligent agent collaboration. *Proc Human Factors Ergon Soc Annu Meet* 60(1):1706–1710. <https://doi.org/10.1177/1541931213601392>
- Stuhlsatz A, Meyer C, Eyben F, Zielke T, Meier G, Schuller B (2011) Deep neural networks for acoustic emotion recognition: raising the benchmarks. *ICASSP IEEE Int Conf Acoust Speech Signal Process Proc*. <https://doi.org/10.1109/ICASSP.2011.5947651>
- Swain M, Routray A, Kabisatpathy P (2018) Databases, features and classifiers for speech emotion recognition: a review. *Int J Speech Technol* 21(1):93–120. <https://doi.org/10.1007/s10772-018-9491-z>
- Sycara K, Sukthankar G (2006) Literature review of teamwork models. Robotics institute, Carnegie Mellon University, 31
- Thomason J, Padmakumar A, Sinapov J, Walker N, Jiang Y, Yedidion H, Hart J, Stone P, Mooney RJ (2020) Jointly improving parsing and perception for natural language commands through human-robot dialog. *J Artif Intell Res* 67:327–374
- Tiferes, J. et al. (2016) The loud surgeon behind the console: understanding team activities during robot-assisted surgery *Journal of Surgical Education* 73:504–512 doi:<https://doi.org/10.1016/j.jsurg.2015.12.009>
- Vergyri, D., Knoth, B., Shriberg, E., Mitra, V., McLaren, M., Ferrer, L., ... Marmar, C. (2015). Speech-based assessment of PTSD in a military population using diverse feature classes. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*.
- Ververidis, D.; Kotropoulos, C. (2003). *A review of Emotional Speech Databases*. 560–574. Retrieved from <http://delab.csd.auth.gr/bci1/Panhellenic/560ververidis.pdf>
- Ververidis D, Kotropoulos C (2006) Emotional speech recognition: resources, features, and methods. *Speech Comm* 48:1162–1181. <https://doi.org/10.1016/j.specom.2006.04.003>
- Vidrascu, L., & Devillers, L. (2005). Real-life emotion representation and detection in call centers data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [https://doi.org/10.1007/11573548\\_95](https://doi.org/10.1007/11573548_95)
- Vogt, T., & André, E. (2005). Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. *IEEE International Conference on Multimedia and Expo, ICME 2005*. <https://doi.org/10.1109/ICME.2005.1521463>
- Vogt, T., André, E., & Wagner, J. (2008). Automatic recognition of emotions from speech: a review of the literature and recommendations for practical realisation. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [https://doi.org/10.1007/978-3-540-85099-1\\_7](https://doi.org/10.1007/978-3-540-85099-1_7)
- Walker GH, Gibson H, Stanton NA, Baber C, Salmon P, Green D (2006) Event analysis of systemic teamwork (EAST): a novel integration of ergonomics methods to analyse C4i activity. *Ergonomics* 49(12–13):1345–1369. <https://doi.org/10.1080/00140130600612846>
- Wallot S, Mitkidis P, McGraw JJ, Roepstorff A (2016) Beyond synchrony: joint action in a complex production task reveals beneficial effects of decreased interpersonal synchrony. *PLoS One* 11(12)
- Wang L, Jamieson GA, Hollands JG (2009) Trust and reliance on an automated combat identification system. *Hum Factors* 51(3):281–291. <https://doi.org/10.1177/0018720809338842>
- Wang Y, Hartman M, Aziz NAA, Arora S, Shi L, Tunison E (2017) A systematic review of the use of LENA technology. *Am Ann Deaf* 162(3):295–311
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*, vol 8. Cambridge University Press, Cambridge
- Watts D, Strogatz S (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442. <https://doi.org/10.1038/30918>
- Weber RO, Aha DW (2003) Intelligent delivery of military lessons learned. *Decis Support Syst* 34(3):287–304
- Wickens CD (2008) Multiple resources and mental workload. *Hum Factors* 50(3):449–455. <https://doi.org/10.1518/001872008X288394>
- Wilson D, Sperber D (2012) *Meaning and relevance*. Cambridge University Press, Cambridge
- Yacoub SM, Simske SJ, Lin X, Burns J (2003) Recognition of emotions in interactive voice response systems. *Interspeech* 50:419–427. <https://doi.org/10.1590/S0036-36342008000500015>
- Yilmaz G (2016) What you do and how you speak matter: behavioral and linguistic determinants of performance in virtual teams. *J Lang Soc Psychol* 35(1):76–97
- Yu, M., Litman, D., & Paletz, S. (2019, May). Investigating the relationship between multi-party linguistic entrainment, team characteristics and the perception of team social outcomes. In *The Thirty-Second International Flairs Conference*.
- Zhang, Z., & Zhang, J. (2006). Driver fatigue detection based intelligent vehicle control. In: *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, Washington, DC, USA, *IEEE Computer Society*, pp. 1262–1265.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.