ARL-TR-8969 ● MAY 2020

DEVCOM
ARMY RESEARCH
LABORATORY

Team Trust in Human–Autonomy Teams: Analysis of Crew Communication during Manned–Unmanned Gunnery Operations

by Anthony L Baker, Sean M Fitzhugh, Daniel E Forster, Ralph W Brewer, Andrea Krausman, Angelique Scharine, and Kristin E Schaefer

NOTICES

Disclaimers

ARL-TR-8969 ● MAY 2020

Team Trust in Human–Autonomy Teams: Analysis of Crew Communication during Manned–Unmanned Gunnery Operations

Anthony L Baker, Sean M Fitzhugh, Daniel E Forster, Andrea Krausman, Angelique Scharine, and Kristin E Schaefer
Human Research and Engineering Directorate, CCDC Army Research Laboratory

Ralph W Brewer
Vehicle Technology Directorate, CCDC Army Research Laboratory

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)* May 2020	**2. REPORT TYPE** Technical Report	**3. DATES COVERED (From - To)** October 2018–May 2020

4. TITLE AND SUBTITLE Team Trust in Human–Autonomy Teams: Analysis of Crew Communication during Manned–Unmanned Gunnery Operations	**5a. CONTRACT NUMBER** W911NF-18-2-0230
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Anthony L Baker, Sean M Fitzhugh, Daniel E Forster, Ralph W Brewer, Andrea Krausman, Angelique Scharine, and Kristin E Schaefer	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) CCDC Army Research Laboratory ATTN: FCDD-RLH-FD Aberdeen Proving Ground, MD 21005-5425	**8. PERFORMING ORGANIZATION REPORT NUMBER** ARL-TR-8969
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES ORCID IDs: Kristin E Schaefer, 0000-0002-1342-3446; Sean M Fitzhugh, 0000-0002-6283-2895; Angelique A Scharine, 0000-0001-8230-1824

14. ABSTRACT

Verbal communication within Soldier and Marine crews was recorded during the Wingman Joint Capability Technology Demonstration's dry-fire and live-fire manned–unmanned teaming gunnery exercises at Fort Benning, Georgia, in October 2018. Their communication was analyzed to identify and assess possible communication-based metrics of team trust and cohesion. Communication flow in five-person crews was aggregated to reveal the proportion of communication for each sender–receiver pair. Findings indicate there were differences in communication patterns between the crews and dry- and live-fire conditions. Aggregate communication flow revealed structural differences in the communication patterns used on-task versus off-task and identified the relative dominance of the commander–gunner communication. Relational event models were used to analyze communication-network dynamics. The models revealed patterns of greater centralization and routinization of communication channels during dry-fire scenarios compared with live-fire scenarios. Language-similarity measures were calculated to evaluate the semantic and syntactic synchrony within the language used by crew members. Measures indicated the Soldier crew exhibited greater variability in its linguistic synchrony, whereas the Marine crew exhibited more stability. Overall, findings suggest these approaches can characterize the interaction and coordination patterns of human–autonomy crews and may be useful for relating communication to team trust and cohesion.

15. SUBJECT TERMS

Wingman, robot, manned–unmanned teaming, communication, analysis, human–autonomy teaming

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Anthony L Baker
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	49	**19b. TELEPHONE NUMBER (Include area code)** 410-278-5396

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

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.

This work supports the US Army Wingman Joint Capability Technology Demonstration (JCTD) and ARL's Human–Autonomy Teaming Essential Research Program, Project 5, on team trust and team cohesion.

The authors would like to thank Debra Patton for assistance with field-data collection during the 2018 Wingman JCTD event. The authors would also like to thank Eric Holder for providing critical guidance and a technical review of this report.

Executive Summary

Effective future human–autonomy teaming (HAT) requires an understanding of team states and processes critical to effective teamwork. As such, communication provides a vehicle for understanding many aspects of teamwork, but it is currently unclear what communication analyses will be most useful to apply to HAT to understand aspects of team coordination, trust, and cohesion, which are three states related to team effectiveness. This report outlines promising approaches to analyzing verbal crew communication as it applies to joint manned–unmanned team gunnery tasks. These approaches were applied to communication data captured during the October 2018 Wingman Joint Capability Technology Demonstration (JCTD) event held at Fort Benning, Georgia.

During the JCTD event, two five-man crews (one Marine crew and one Soldier crew) conducted live-fire gunnery evaluation exercises[1] using the Wingman manned–unmanned vehicle platforms. For this event, the Wingman platform consisted of two High-Mobility Multipurpose Wheeled Vehicles: a command-and-control vehicle and a robotic combat vehicle. The command-and-control vehicle carried the crew while the robotic combat vehicle was unmanned and weaponized. For this event, the crew was able to control the vehicle mobility through either teleoperation or mobility autonomy; however, teleoperation controls were only available for the weapon system.

To improve the effectiveness of future human–autonomy teams, Project 5 of the US Army Combat Capabilities Development Command Army Research Laboratory's HAT Essential Research Program seeks to evaluate and assess metrics of trust and team cohesion that are applicable to human–autonomy teams. There are many ways to evaluate team trust and cohesion, and this report outlines our efforts to identify novel, communication-based metrics for those constructs. Three approaches are described: aggregate communication-flow modeling, relational event modeling, and linguistic similarity. These methods were selected on the basis of their potential ability to describe team trust and cohesion in future human–autonomy teams.

Findings suggest task conditions affected crew communication in various ways. Aggregate communication-flow data revealed differences in the communication patterns of crews during on-task versus off-task phases, and highlighted the importance of the vehicle commander–gunner pairing during gunnery operations.

---

[1] [DOA] Headquarters, Department of the Army. Training and qualification, crew. Washington (DC): Headquarters, Department of the Army (US); 2015 Mar. Training Circular No.: TC 3-20.31.

Relational event models demonstrated more decentralized and routinized communication during live-fire exercises; additionally, intrateam shifts in communication suggest differences between how teams utilize roles. Linguistic-similarity analyses revealed the appearance of effortful knowledge building among dyads—those exhibiting greater levels of synchrony were also communicating more fluidly about their respective roles, challenges, and preferences during the task. Our findings suggest a team's communication patterns can provide a window into changes in team trust and cohesion, in support of the objectives of HAT Project 5. Follow-up work will involve development and refinement of these methods, as well as further investigation into how these communication metrics can be linked with other data from human–autonomy teams.

1. Introduction

The development of advanced autonomy-enabled systems is expected to shape the future battlefield. The Army anticipates that autonomy will serve as a force multiplier, extending the ability of Warfighters to impact the battlespace, reducing the number of Warfighters in harm's way, and providing new ways of operating that should improve the flexibility and capabilities of teams (ARCIC 2017). Much ongoing research is targeted at designing adaptive, intelligent systems that can work fluidly and naturally with human team members to achieve the mission (Barnes et al. 2017; DeCostanza et al. 2018, p. 38; Marathe et al. 2018).

While future human–autonomy systems and concepts such as the Next Generation Combat Vehicle seek to integrate advanced technologies and autonomous systems within teams of Soldiers, it is unclear how trust and cohesion, both essential for team function and performance, will be affected by these novel team configurations. The future battlefield demands team structures that can adapt and reconfigure to achieve the mission, but it is unknown how trust and cohesion will be shaped by these demands, given that teams may not have the opportunity to develop trust and cohesion through normal means. If human–autonomy teams are to be successful, it is critical to understand how team trust and cohesion can be developed and maintained in the context of human–autonomy teams.

In human–autonomy teaming (HAT), the autonomy introduces unique capabilities and challenges into a team, and much of that challenge stems from how team communication is affected by autonomy. To build more advanced systems and better human–autonomy teams, it is fundamental that we understand how team communication affects, and is affected by, autonomy. Team communication is the vehicle through which teams exchange information, coordinate toward goals, and solve problems. Therefore, with a better understanding of team communication, we may be able to better understand how trust and cohesion evolve throughout the course of a team's interactions, as well as how those interactions result in team outcomes.

1.1 Background: Team Communication

Communication shapes critical aspects of teamwork, performance, cohesion, and trust. Within teams, communication is critical to maintaining a common understanding (Clark and Brennan 1991) as it is the vehicle through which team members share information, align toward goals, and resolve disagreements (Salas et al. 2005). But when teams fail to communicate effectively, they can fail to work together effectively. For example, approximately 40% of the situational factors in naval groundings were accounted for by poor communication (Macrae 2009), and approximately 70%–80% of all aviation accidents from 1980 to 2000 could be attributed to communication errors (Sexton and Helmreich 2000). Miscommunication was further cited as an important factor in two US Army Black Hawk helicopters being shot down in a friendly fire incident during Operation Provide Comfort in 1994 (Snook 2000).

By identifying patterns of team communication, it becomes possible to better understand and evaluate team performance. These patterns reveal how a team works together to complete tasks, distribute information, and coordinate goal-directed behaviors, offering a window into the team's successful and unsuccessful coordination processes (Kiekel et al. 2001; Sacks et al. 1974; Tiferes et al. 2016). However, with the increasing complexity of intelligent agents and autonomous systems, better methods and metrics are needed for understanding team communication in HAT. Therefore, we posit that team communication patterns offer useful data that, when analyzed, can provide key insights into team coordination, cohesion, and trust.

1.2 Use Case: Wingman Joint Capability Technology Demonstration (JCTD)

The purpose of the US Army's JCTD program was to develop advanced robotic technologies and assess increased autonomous capabilities for joint manned and unmanned ground combat vehicles. Wingman is the first example of a real-world prototype human–autonomy team that was designed for gunnery operations. While the current vehicles are a Stryker command and control (C2) vehicle for crew operations and a M113 robotic combat vehicle (RCV), the experimentation reported here used the earlier prototype High-Mobility Multipurpose Wheeled Vehicle (HMMWV) in C2 and RCV variants. The design of the C2 HMMWV supported a five-man crew comprising the manned vehicle's driver, vehicle commander (VC), and Long-Range Advanced Scout Surveillance System (LRAS3) operator and robotic-vehicle operator (RVO) and robotic-vehicle gunner (RVG), paired with a single unmanned weaponized robotic ground vehicle (see Fig. 1). Within the context of the gunnery tasks conducted by the crews during the Wingman JCTD event, the VC served as the crew leader, the LRAS3 operator acquired target ranges and relayed range information to the crew, the RVG operated the robotic weapon system, the RVO was tasked with mobility of the robotic vehicle, and the driver moved the C2 vehicle between locations as needed. Crews communicated verbally over an intercom and monitored and controlled their weapon and mobility autonomy systems via a computer interface. For this report, our focus was on the verbal communication used within each crew as they completed their gunnery exercises using the C2 vehicle and RCV.

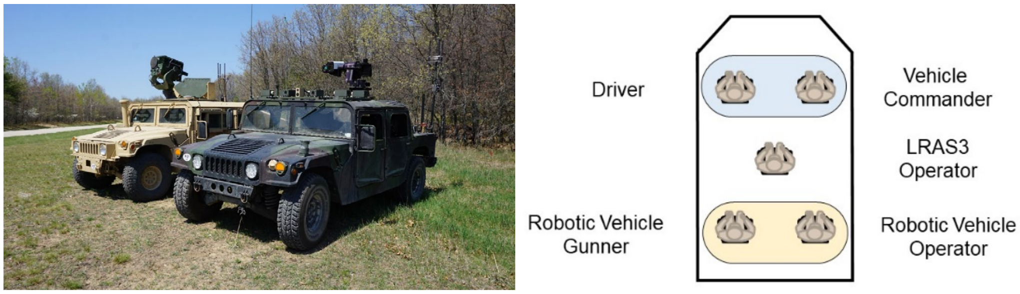


Fig. 1 Image on left depicts the C2 (left) and the RCV (right) HMMWV prototype platforms; image on right is a schematic of crew layout onboard the C2 vehicle

1.3 Study Objectives

The goal of the Human–Autonomy Teaming Essential Research Program Project 5 is to develop effective metrics of team trust and team cohesion that can be applied to human–autonomy teams. Because communication is the vehicle through which many team states and processes emerge, communication analyses can provide insights into the trust, cohesion, and performance of human–autonomy teams. In this report, methods are detailed and discussed for assessing the structure and content of verbal communication used among crew members in the Wingman JCTD gunnery event, along with how findings can provide insights into the trust, cohesion, and performance of those crews as well as human–autonomy teams in general. Therefore, the objectives are as follows:

- Analyze verbal crew communication in a field exercise involving gunnery operations in a human–autonomy team
- Evaluate how different aspects of communication, such as structure, content, and sequencing, relate to the crew's performance
- Discuss how analytical methods can provide insights into the trust and cohesion of human–autonomy teams

These analyses are presented as a proof-of-concept for deriving team trust and cohesion insights from the communication of a human–autonomy crew. The conclusions lay the groundwork for additional development and testing of these methods for understanding how trust and cohesion can be derived from a crew's communication. Follow-up work will leverage new data sets to test these approaches and evaluate their ability to predict trust and cohesion in human–autonomy crews. Additional efforts will seek to understand how these communication metrics relate to other metrics of performance, trust, and cohesion in human–autonomy teams.

2. Methods

2.1 Participants

Two crews participated in this study: one crew comprised Marines from the 1st Marine Logistics Group Combat Skills Training School (CSTS) and the second crew comprised Soldiers from the 3rd Squadron, 16th Cavalry Regiment. An engineering team member filled the role of the C2 vehicle driver for both crews. The Warfighters filled the remaining four crew stations.

The selection of the CSTS Marines for the Wingman JCTD was based on their experience and expertise in machine-gun team operations, concept of employment, and employing tactics, techniques, and procedures (TTPs). As infantry subject-matter experts, the CSTS Marines train Marines and Sailors in regiments, battalions, and companies across the Marine Logistics Group. Their primary focus is on combat training courses. These courses cover tactical leadership principles, machine-gun functions, combat orders, and procedures to counter threats and mitigate risks to Marine forces conducting tactical convoys. Each of the four Marines who attended the JCTD has deployed in support of combat operations during Operation Enduring Freedom/Operation Iraqi Freedom.

The four Soldiers selected from the 16th Cavalry Regiment hold the military occupational specialty of 19D, Cavalry Scout. They are proficient in reconnaissance, surveillance, and target acquisition, and are subject-matter experts in movement and maneuver. These Soldiers belong to the 316th Cavalry Brigade, whose primary mission is to train Cavalry enlisted soldiers (19D and 19K, One Station Unit Training) and conduct other training at Fort Benning to include the Armor Officer Basic Course, Army Reconnaissance Course, and Cavalry Leader's Course. As such, these Soldiers have a premiere pedigree to understand the TTPs necessary to fight and win on today's technologically advanced battlefield, especially the ability to move, shoot, and communicate. All of the Soldiers selected had previously qualified on a crew gunnery evaluation Table VI, making them familiar with the requirements from Training Circular (TC) 3-20.21 on which the Wingman event based its gunnery operations. The Marine crewmen had prior experience working together, as did the Army crewmen.

2.2 Schedule of Events

The gunnery events included the following:

- Monday, 22 October 2018: Classroom training, simulation training, and one dry-fire training exercise for each crew
- Tuesday, 23 October 2018: Observed engineering team's live-fire demonstration; Soldier and Marine crews each conducted one stationary live-fire exercise
- Wednesday, 24 October 2018: Soldier and Marine crews each conducted one blank-fire exercise and two dry-fire exercises
- Thursday, 25 October 2018: Soldier and Marine crews each conducted one dry-fire exercise and one stationary live-fire exercise

For a full review of this event, training, and additional feedback and analysis, see Schaefer et al. (2019b).

2.2.1 Dry-Fire and Blank-Fire Engagement Runs

In a crawl–walk–run approach, the crews began by conducting evaluation exercises on the course without using any ammunition. The crews went through the motions of engaging the targets using the DIDEA process (detect, identify, decide, engage, assess) and dry-firing the gun. This was important to synchronize their crew duties prior to adding ammunition to the process. Next, crews conducted the same evaluation exercises on the course using blank ammunition to practice engaging targets and firing the weapon without any rounds going down range. The blank-fire runs provided crews with the additional feedback of hearing the weapon fire while performing their training.

2.2.2 Stationary Live-Fire Engagement Runs

Once the crews were sufficiently trained and evaluated with dry and blank fire, they were allowed to take part in the stationary live-fire engagement run. The stationary live-fire engagements were similar to the dry- and blank-fire engagement runs with two differences. First, the crews used live ammunition. Second, to comply with an Army Test and Evaluation Command safety memo restricting live fire while moving, the RCV was chocked to prevent vehicle movement while the weapon was loaded. Defensive engagements did not change, but offensive engagements were changed to reflect traffic-control-point procedures and scoring. A traffic-control-point engagement treats the evaluation as an offensive one with regard to timing, but the vehicle does not move.

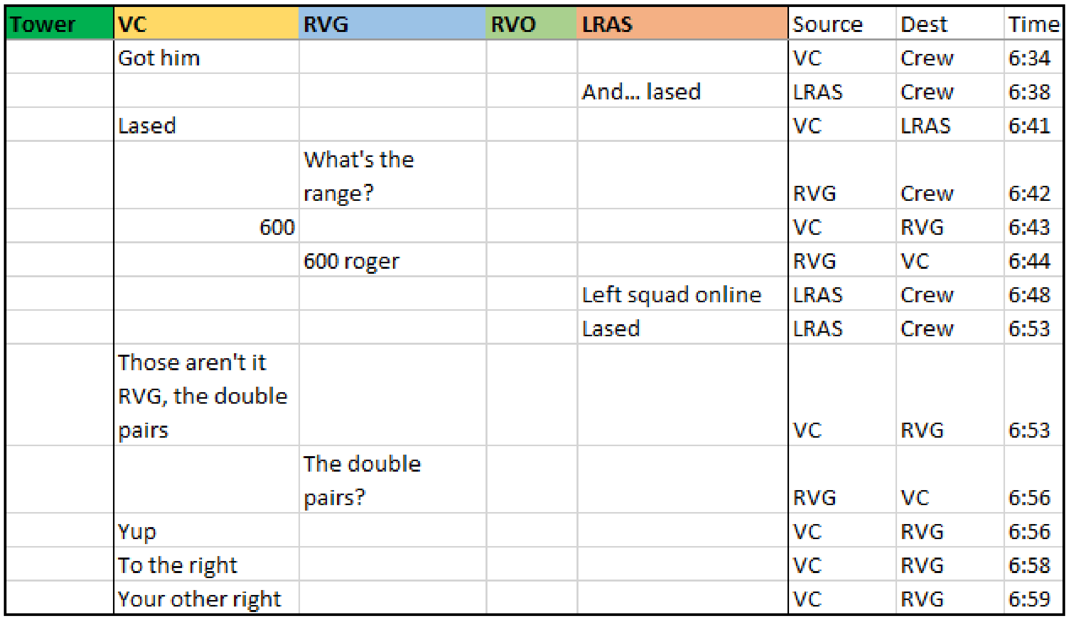
2.3 Equipment

Team communication was transmitted within the crew via standard Combat Vehicle Crewmen headsets used in the Army. The crew's communication was all on the same network, such that any crew member could speak to any other and all crew members could hear all communication. This stream of communication was sent to the Tower, a centralized location in the field exercise that oversaw the crew's activities. In the Tower, the communication stream was played live on a laptop, which recorded all communication using a screen record function.

2.4 Analytical Plan for Crew Communication

One dry-fire run and one live-fire run were analyzed for each crew, totaling four communication data sets. The specific runs used for analyses were selected because they occurred on the same day (Thursday). Selecting only live and dry runs that were conducted on the same day reduced the possibility that any differences between the live and dry data could be due to learning effects. It was also important to select two sets of runs for each crew to evaluate how crew communication patterns might be different between dry and live runs.

Both audio and transcriptions of the audio data were used in the following analyses. Excel files were used to transcribe the data during all seven target engagements for each of the dry- and live-fire events. For this data set, nonfluencies and nonverbal expressions (e.g., "huh" or "um") were excluded from transcription. Each spoken communication event was coded for its source (the speaker) and destination (the intended recipient) and the timestamp based on the communication flow paradigm (see Baker et al. 2019; Schaefer et al. 2019a). If a communication event was not directed at a specific crew member, the destination was listed as "Crew". Crew was also listed as the destination if the intended recipient of a communication event was not decisively clear based on context. A sample of a crew communication transcript is provided in Fig. 2.



Tower	VC	RVG	RVO	LRAS	Source	Dest	Time
	Got him				VC	Crew	6:34
				And... lased	LRAS	Crew	6:38
	Lased				VC	LRAS	6:41
		What's the range?			RVG	Crew	6:42
	600				VC	RVG	6:43
		600 roger			RVG	VC	6:44
				Left squad online	LRAS	Crew	6:48
				Lased	LRAS	Crew	6:53
	Those aren't it RVG, the double pairs				VC	RVG	6:53
		The double pairs?			RVG	VC	6:56
	Yup				VC	RVG	6:56
	To the right				VC	RVG	6:58
	Your other right				VC	RVG	6:59

Fig. 2 Sample of crew-communication transcript; each communication event is on a separate line

For the purpose of this report, the communication analyses included aggregate communication flow, relational event modeling, and linguistic similarity. The following methods were selected based on their potential applicability in deriving trust and team cohesion from team communication. For a more complete list of additional communication analysis methods in consideration, see Baker et al. (2019).

2.4.1 Communication Flow

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 who speaks to whom throughout a team interaction, can allow one to evaluate team processes at the cognitive and the interpersonal level (Fischer et al. 2007). Communication flow can be represented visually 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.

As an example of how communication flow can be modeled to represent how a team performs its tasks, Fischer et al. (2007) noted how often team members in a simulated search-and-rescue task responded to each other’s communication, as depicted in the flow diagrams in Fig. 3. These show structural differences in the communication patterns of successful and unsuccessful teams, revealing that successful search-and-rescue teams had a more equal distribution of communication, whereas unsuccessful teams tended to involve a few team members dominating the discourse.

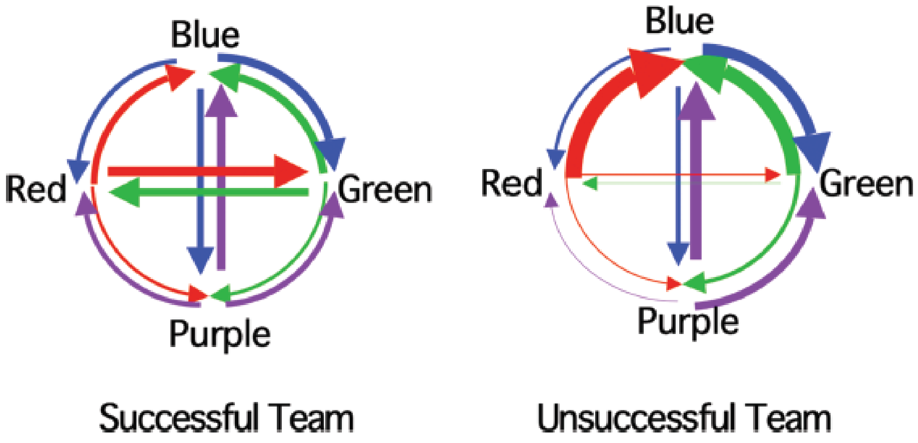


Fig. 3 Flow diagrams for successful and unsuccessful teams (Fischer et al. 2007)

For this report, communication flow was assessed in the same vein as Fischer et al. (2007), but this approach differs from theirs by identifying, rather than ignoring, communication that was not targeted at a specific team member. For example, a crew member may state "Hey team, I am moving to Point B". By accounting for this nontargeted communication to the crew, it becomes easier to capture information that team members might share with their team as a whole to improve situation awareness and update shared mental models (Cannon-Bowers et al. 1993; Wildman et al. 2014).

2.4.2 Relational Event Modeling

In a second approach, a relational event model was built from the transcribed data to analyze communication network dynamics within each team. Rather than treating the network as a time-aggregated representation of the structure of interactions, the network was represented as a timestamped series of discrete, directed interactions called relational events (Butts 2008). By examining network dynamics in this framework, it is possible to capture time-ordered features of interactions such as message-forwarding. In addition, the temporal context of interactions was captured by preserving the exact timing and order of messages. Each relational event (a) transpires from a sender (i) to a receiver (j) at a given time (t). While the nature of the interaction may vary—sending messages, economic transactions, dominance contests, spreading disease—in this case we use the spoken interactions within the crew. However, we only examine the structure of interactions as this method does not take content into account (although one could conceivably filter the data to examine dynamics of content of interest such as information provision or conflict). A given relational event (*a*) belongs to a larger event history (A_t) and evaluates the likelihood of observing that event history with the relational event model.

Building on a framework for traditional event-history models, the relational event model posits the joint likelihood of event history (A_t), which is simply the product of the likelihoods of each communication act (*a*). The conditional likelihood for the *i*th event in A_t is equal to the hazard (following terminology used in the survival analysis branch of statistics) for relational event a_i. 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 toward or against specific types of communication acts.

In practice, relational event models demonstrate how communication rates are affected by sender attributes (e.g., perceived team trust and experience in one's role), receiver attributes (e.g., perceived team cohesion and individual performance) communication attributes (e.g., communication within a vehicle vs. communication across vehicles), and network statistics (e.g., centrality, tie weight, and structural cohesion). This gives us two avenues for examining trust and cohesion within the team: 1) using terms in the model to evaluate effects of structural cohesion and perceived trust or cohesion on communication rates and 2) relating model coefficients to team-level trust or cohesion to identify whether particular patterns of communication (e.g., more centralized, more reliant on stronger ties) are associated with our team states of interest.

2.4.3 Linguistic Similarity

Linguistic Style Matching (LSM; Niederhoffer and Pennebaker 2002), Latent Semantic Similarity (LSS; Babcock et al. 2014), and ConversAtion level Syntax SImilarity Metric (CASSIM; Boghrati et al. 2018) were computed to examine linguistic similarity among team members. Scholars have used linguistic-similarity metrics to study team processes in human teams (Gonzales et al. 2010), naturally making it a focus in human–autonomy teams, as well (Demir et al. 2017). In the context of human–autonomy teams, linguistic-similarity metrics could help track synchrony among humans and between humans and autonomy, and even be used to inform how autonomous agents should adapt their communication styles to humans.

LSM is defined as the proportion of so-called “function” words (e.g., pronouns, adverbs, and prepositions) in a conversation. It is a dictionary-based lexical metric, meaning it requires matching words to an existing dictionary to compute levels of synchrony, while ignoring structure. The logic underlying LSM is based on the understanding that when people have established a subject of conversation (e.g., “our mutual friend John from the city”), their conversation will be dominated by words that reference the subject without specifying details (e.g., him from there) (Niederhoffer and Pennebaker 2002). If speakers exhibit similar rates of using content-free words, then we can infer the speakers are forming common knowledge at similar rates. LSM is based on proportions of words, thereby restricting the range of values between 0 (no synchrony) and 1 (perfect synchrony). Calculating similarity across speakers is conducted on a pairwise basis—to understand linguistic synchrony among team members, therefore, we averaged across pairwise values.

LSS is a semantic metric, but it does not match words to existing word categories—rather, it evaluates word clusters using latent semantic analysis (LSA) to then assess whether people are contributing similarly to the conversation. By examining word clusters, it is possible to understand the themes of verbal communication (Landauer et al. 1998). To extend this analysis to a conversation with two or more speakers, scholars examine the similarity of these LSA-derived word clusters across speakers. This metric ranges from –1 (perfect asynchrony) to 1 (perfect synchrony). As with LSM, LSS is computed at a pairwise level, so we averaged across pairwise similarity to understand team-level synchrony.

In contrast, CASSIM is a syntactic metric, meaning it requires extracting the structure of words in a sentence to compute levels of synchrony while ignoring word meanings. Unlike LSM and LSS, it does not compute similarity based on the types of words used; rather, CASSIM quantifies the structural similarity of statements in a conversation (Boghrati et al. 2018). To do so, each sentence is deconstructed under a “constituency parse tree”, which branches components of the overall sentence into nouns, verb phrases, noun phrases, and other structural components. By default, CASSIM uses a parsing algorithm developed by Klein and Manning (2003), which has the desirable properties of being relatively efficient and accurate, but CASSIM is equipped to handle other methods of parsing (making it forward-compatible as parsers become more advanced). Once a parse tree has been determined for two sentences, one could compute the minimum number of edits required to equate the two parse trees. With this *Edit Distance* algorithm (Navarro 2001), CASSIM can estimate the similarity between the sentences of two documents. Although CASSIM is equipped to compute similarity using multiple documents, unlike standard LSM and LSS packages, CASSIM appears to do so by averaging across pairwise values (as we did manually with LSM and LSS).

Linguistic synchrony was analyzed at both team and dyadic levels to account for synchronization across the entire team and across dyads within the team. We suspect synchronization may provide insight into cohesion within the team. However, linguistic-synchrony metrics as predictors of performance may require more nuance—after all, cohesive groups who readily come to agreements may fall to the perils of "groupthink", a phenomenon in which consensus becomes a goal in and of itself, irrespective of whether such consensus is the result of mutual recognition of an optimal choice (Janis 1972; Mullen et al. 1994). In any case, more work and more data are needed to understand how synchrony and asynchrony map onto team cohesion, especially in the context of military human–autonomy teams. Thus, the present work details an initial step in understanding whether linguistic synchrony within a manned–unmanned gunnery crew can provide insight into their trust and cohesion.

As the social networks analysis indicates, some dyads of a team may never communicate whereas others will almost exclusively communicate. At the team level, linguistic synchrony was analyzed for each service's team (Soldier vs. Marine) during each run (live fire vs. dry fire) and each engagement (one through seven). At the dyadic level, linguistic synchrony was computed for each crew during each run, but engagement-level synchrony was ignored because it was rare to find an engagement during which there was enough back-and-forth talk for a given dyad.

3. Results

A trained vehicle crew evaluator assessed the gunnery crew during its runs using Table VI scoring criteria. Each target engagement trial (TET) was worth a maximum of 100 points, and each run consisted of 7 TETs. Crew performance scores are provided in Table 1. These scores are determined by the difficulty of the TET (target type, target range, etc.), the type of engagement (i.e., offensive or defensive), the number of targets, and any penalties (e.g., communication errors). The scores are calculated using the guidelines in TC 3-20.21 and recorded on a Department of the Army Form 8265.

Table 1 Crew qualification scores

Day run	Crew	TET 1	TET 2	TET 3	TET 4	TET 5	TET 6	TET 7	Day total	Qualified # TET	Qualified? 70%
Dry	Marine	94	89	0	98	98	100	100	579	6	Yes
Dry	Army	97	89	100	99	92	100	100	677	7	Yes
Live	Marine	50	47	0	50	0	80	100	327	2	No
Live	Army	50	53	100	73	87	100	100	563	4	Yes

Note: All TET scores are presented out of a maximum score of 100. Over 70 is considered a passing score for the TET, and scores above 70 are highlighted in green. A total score over 70% is required to qualify on the Table VI gunnery evaluation for each run.

From Table 1, we see that Army and Marine teams were approximately similar in performance during the Dry runs, but crews performed very differently on Live runs. This difference is partly accounted for by the differences in task demands between Live and Dry runs. The Live runs were the only runs that involved actual ammunition hitting a target, whereas for the Dry runs crews were credited with a "target hit" if they slewed the weapon system over the target and the gunner verbally indicated firing. Therefore, performance on the Live runs reflected both crew coordination and gunnery marksmanship. The Army crew in this study was familiar with gunnery exercises for human teams, so its performance on the live runs may have reflected this experience.

While performance scores provide some insights into the differences between the crews, more information is needed to understand why these outcomes occurred. It is not enough to only consider performance scores, given that they only provide limited insight into what happened during the interaction. To better understand what happened during the gunnery engagements, and to gain insight into possible aspects of their trust and cohesion, we will assess the crew's communication using several analytical approaches discussed in the following sections. In other words, the following analyses help reveal how the teams achieved their respective performance scores.

3.1 Communication Flow

To produce usable communication flow data, each communication event in the Fort Benning 2018 audio transcripts was coded for its sender (i.e., the speaker) and receiver (i.e., the intended recipient of the communication). After this, the amount of communication events accounted for by each sender and each sender–receiver pair was aggregated and evaluated for trends in communication patterns between and within crews. Communication was broken down across both crews (Army and Marines) and across both gunnery conditions (dry fire and live fire). Further, within the gunnery conditions, crews were not always actively engaging targets; between the seven engagements, crews were often either moving their vehicles to the next position or waiting for an engagement to begin. Therefore, communication during all runs was further separated into *engagements* (i.e., communication during target engagements) and *interengagement* (i.e., communication outside any target engagement). Tables 2 and 3 provide data about the senders of communication events across all conditions and crews. Table 2 specifically provides the number of communication events initiated by each crew member. This is intended to highlight the extent to which each crew member spoke. In other words, it provides a clear way to see which crew members initiated more communication events. Table 3 provides the same information as Table 2, except Table 3 provides the data as a proportion. In other words, Table 3 demonstrates the percent of a crew's communication events that was initiated by each crew member. For example, in Table 3 we see the VC initiated approximately 45% of all communication events in both the Army and Marine crews during the dry runs, whereas that count dropped to approximately 36% for the live runs.

Table 2 Number of communication events initiated by each crew member

			VC	**RVG**	**RVO**	**LRAS**	**Tower**	**Driver**
Dry-Fire Run	Marines	**All Comms**	120	59	18	37	20	14
		Engagements	44	29	2	25	0	0
		Interengagement	76	30	16	12	20	14
	Army	**All Comms**	85	38	26	11	19	8
		Engagements	27	30	0	10	0	0
		Interengagement	58	8	26	1	19	8
Live-Fire Run	Marines	**All Comms**	153	136	28	54	30	24
		Engagements	56	54	1	24	3	0
		Interengagement	97	82	27	30	27	24
	Army	**All Comms**	139	80	45	63	32	22
		Engagements	64	39	0	30	4	0
		Interengagement	75	41	45	33	28	22

Note: "All Comms" is the sum of events across engagements and interengagement.

Table 3 **Proportion of communication events initiated by each crew member**

			VC	RVG	RVO	LRAS	Tower	Driver
Dry Runs	Marines	**All Comms**	44.80%	22.00%	6.70%	13.80%	7.50%	5.20%
		Engagements	44.00%	29.00%	2.00%	25.00%	0.00%	0.00%
		Interengagement	45.20%	17.90%	9.50%	7.10%	11.90%	8.30%
	Army	**All Comms**	45.50%	20.30%	13.90%	5.90%	10.20%	4.30%
		Engagements	40.30%	44.80%	0.00%	14.90%	0.00%	0.00%
		Interengagement	48.30%	6.70%	21.70%	0.80%	15.80%	6.70%
Live Runs	Marines	**All Comms**	36.00%	32.00%	6.60%	12.70%	7.10%	5.60%
		Engagements	40.60%	39.10%	0.70%	17.40%	2.20%	0.00%
		Interengagement	33.80%	28.60%	9.40%	10.50%	9.40%	8.40%
	Army	**All Comms**	36.50%	21.00%	11.80%	16.50%	8.40%	5.80%
		Engagements	46.70%	28.50%	0.00%	21.90%	2.90%	0.00%
		Interengagement	30.70%	16.80%	18.40%	13.50%	11.50%	0.090

Several trends are immediately apparent. First, the VC is the crew member who often initiated the largest number of communication events. This is expected due to the dynamics outlined in TC 3-20.31 (DOA 2015). The VC serves a role as the leader of the crew who must orchestrate crew members and coordinate gunnery information between the LRAS3 operator and the gunner to successfully engage targets. Therefore, it is not surprising the VC was the top initiator of communication events in most of the scenarios. Second, the RVG was generally the second-most common initiator of communication events. Because gunners often needed to relay information about target acquisition, engagement, and sensing (i.e., noticing where rounds went after firing), this provides support for the idea that within this gunnery task, the VC–RVG pair is integral to the team's performance. Third, the relative distribution of crew communication between *engagements* and *interengagement* appears to differ. Tables 4 and 5 represent the distribution of communication events accounted for by the various possible sender–receiver pairs in each condition and crew. Table 4 contains data pertaining to the dry run, and Table 5 contains data for the live run. Both tables represent the sender–receiver pairs in descending order for each task phase (i.e., engagements, interengagement, or all communication) and crew (Army or Marine). For example, Table 4 (dry run) demonstrates that the greatest proportion of messages among the Army crew members during target engagements was sent from the VC to the RVG; in other words, 37.3% of all of the crew's communication during target engagements was accounted for by the VC speaking to the RVG.

Table 4 Live runs

Marines						Army					
All comms		Engagements		Downtime		All comms		Engagements		Downtime	
VC–RVG	0.228	VC–RVG	0.326	VC–RVG	0.181	VC–RVG	0.202	VC–RVG	0.416	RVO–Crew	0.148
RVG–Crew	0.148	RVG–Crew	0.261	RVG–VC	0.122	RVG–Crew	0.136	RVG–Crew	0.204	VC–Tower	0.098
RVG–VC	0.118	LRAS3–Crew	0.130	RVG–Crew	0.094	RVO–Crew	0.094	LRAS3–RVG	0.131	RVG–Crew	0.098
LRAS3–Crew	0.066	RVG–VC	0.109	VC–Tower	0.080	LRAS3–RVG	0.092	LRAS3–Crew	0.088	Tower–Crew	0.098
Tower–Crew	0.059	VC–LRAS3	0.058	Tower–Crew	0.077	Tower–Crew	0.073	RVG–VC	0.073	VC–RVG	0.082
VC–Tower	0.054	LRAS3–RVG	0.036	RVO–Crew	0.059	VC–Tower	0.066	VC–Crew	0.044	LRAS3–RVG	0.070
RVO–Crew	0.040	VC–Crew	0.022	Driver–RVO	0.042	LRAS3–Crew	0.055	Tower–Crew	0.029	VC–Crew	0.057
LRAS3–RVG	0.038	RVG–LRAS3	0.022	RVG–LRAS3	0.038	VC–Crew	0.052	VC–Tower	0.007	VC–RVO	0.049
RVG–LRAS3	0.033	Tower–Crew	0.022	LRAS3–RVG	0.038	RVG–VC	0.042	RVG–LRAS3	0.007	Driver–RVO	0.041
VC–Crew	0.031	RVO–VC	0.007	VC–Crew	0.035	VC–RVO	0.031	VC-RVO	0.000	LRAS3–Crew	0.037
VC–LRAS3	0.028	LRAS3–VC	0.007	LRAS3–Crew	0.035	Driver–RVO	0.026	VC-LRAS	0.000	Driver–Crew	0.029
Driver–RVO	0.028	VC-RVO	0.000	VC–RVO	0.028	Driver–Crew	0.018	VC-Driver	0.000	RVG–VC	0.025
VC–RVO	0.019	VC-Tower	0.000	RVG–Driver	0.024	RVG–LRAS3	0.016	RVG-RVO	0.000	RVG–LRAS3	0.020
RVG–Driver	0.016	VC-Driver	0.000	Driver–Crew	0.021	RVG–Tower	0.013	RVG-Tower	0.000	RVG–Tower	0.020
RVO–VC	0.014	RVG-RVO	0.000	RVO–VC	0.017	LRAS3–RVO	0.013	RVG-Driver	0.000	LRAS3–RVO	0.020
LRAS3–VC	0.014	RVG-Tower	0.000	LRAS3–VC	0.017	RVO–Driver	0.010	RVO-VC	0.000	RVO–Driver	0.016
Driver–Crew	0.014	RVG-Driver	0.000	Tower–VC	0.017	VC–Driver	0.008	RVO-RVG	0.000	VC–Driver	0.012
Tower–VC	0.012	RVO-RVG	0.000	VC–LRAS3	0.014	RVO–VC	0.008	RVO-LRAS	0.000	RVO–VC	0.012
RVO–Driver	0.009	RVO-LRAS	0.000	RVO–Driver	0.014	Tower–VC	0.008	RVO-Crew	0.000	Tower–VC	0.012
LRAS3–Driver	0.007	RVO-Crew	0.000	LRAS3–Driver	0.010	Driver–VC	0.008	RVO-Tower	0.000	Driver–VC	0.012
Driver–RVG	0.007	RVO-Tower	0.000	Driver–RVG	0.010	VC–LRAS3	0.005	RVO-Driver	0.000	VC–LRAS3	0.008
Driver–LRAS3	0.005	RVO-Driver	0.000	Driver–LRAS3	0.007	RVO–Tower	0.005	LRAS-VC	0.000	RVO–Tower	0.008
RVG–RVO	0.002	LRAS-RVO	0.000	RVG–RVO	0.003	LRAS3–VC	0.005	LRAS-RVO	0.000	LRAS3–VC	0.008
RVG–Tower	0.002	LRAS-Tower	0.000	RVG–Tower	0.003	RVG–RVO	0.003	LRAS-Tower	0.000	RVG–RVO	0.004
RVO–LRAS3	0.002	LRAS-Driver	0.000	RVO–LRAS3	0.003	Tower–RVG	0.003	LRAS-Driver	0.000	Tower–RVG	0.004
LRAS3–RVO	0.002	Tower-VC	0.000	LRAS3–RVO	0.003	Driver–RVG	0.003	Tower-VC	0.000	Driver–RVG	0.004
Driver–VC	0.002	Tower-RVG	0.000	Driver–VC	0.003	Driver–Tower	0.003	Tower-RVG	0.000	Driver–Tower	0.004
VC-Driver	0.000	Tower-RVO	0.000	VC-Driver	0.000	RVG-Driver	0.000	Tower-RVO	0.000	RVG-Driver	0.000
RVO-RVG	0.000	Tower-LRAS	0.000	RVO-RVG	0.000	RVO-RVG	0.000	Tower-LRAS	0.000	RVO-RVG	0.000
RVO-Tower	0.000	Tower-Driver	0.000	RVO-Tower	0.000	RVO-LRAS	0.000	Tower-Driver	0.000	RVO-LRAS	0.000
LRAS-Tower	0.000	Driver-VC	0.000	LRAS-Tower	0.000	LRAS-Tower	0.000	Driver-VC	0.000	LRAS-Tower	0.000
Tower-RVG	0.000	Driver-RVG	0.000	Tower-RVG	0.000	LRAS-Driver	0.000	Driver-RVG	0.000	LRAS-Driver	0.000
Tower-RVO	0.000	Driver-RVO	0.000	Tower-RVO	0.000	Tower-RVO	0.000	Driver-RVO	0.000	Tower-RVO	0.000
Tower-LRAS	0.000	Driver-LRAS	0.000	Tower-LRAS	0.000	Tower-LRAS	0.000	Driver-LRAS	0.000	Tower-LRAS	0.000
Tower-Driver	0.000	Driver-Crew	0.000	Tower-Driver	0.000	Tower-Driver	0.000	Driver-Crew	0.000	Tower-Driver	0.000
Driver-Tower	0.000	Driver-Tower	0.000	Driver-Tower	0.000	Driver-LRAS	0.000	Driver-Tower	0.000	Driver-LRAS	0.000

Table 5 Dry runs

Marines						Army					
All comms		Engagements		Downtime		All comms		Engagements		Downtime	
VC–RVG	0.201	VC–RVG	0.280	VC–RVG	0.155	VC–RVG	0.182	VC–RVG	0.373	VC–Tower	0.167
RVG–Crew	0.108	RVG–Crew	0.180	VC–Tower	0.155	RVG–Crew	0.128	RVG–Crew	0.313	VC–RVO	0.158
RVG–VC	0.104	LRAS3–Crew	0.180	RVG–VC	0.101	VC–RVO	0.112	LRAS3–Crew	0.149	Tower–VC	0.133
VC–Tower	0.097	RVG–VC	0.110	Tower–VC	0.101	VC–Tower	0.107	RVG–VC	0.134	RVO–VC	0.125
LRAS3–Crew	0.086	VC–LRAS3	0.100	RVG–Crew	0.065	Tower–VC	0.086	VC–RVO	0.030	VC–RVG	0.075
Tower–VC	0.063	VC–Crew	0.050	VC–RVO	0.054	RVO–VC	0.080	VC-LRAS	0.000	RVO–Crew	0.075
VC–LRAS3	0.056	LRAS3–VC	0.050	RVO–Crew	0.048	RVG–VC	0.059	VC-Crew	0.000	VC–Driver	0.042
VC–Crew	0.041	LRAS3–RVG	0.020	VC–Crew	0.036	LRAS3–Crew	0.059	VC-Tower	0.000	Driver–VC	0.042
VC–RVO	0.037	VC–RVO	0.010	Driver–Crew	0.036	RVO–Crew	0.048	VC-Driver	0.000	VC–Crew	0.033
RVO–Crew	0.030	RVO–VC	0.010	VC–LRAS3	0.030	VC–Driver	0.027	RVG-RVO	0.000	RVG–Crew	0.025
LRAS3–VC	0.030	RVO–RVG	0.010	RVO–VC	0.030	Driver–VC	0.027	RVG-LRAS	0.000	RVG–Tower	0.025
RVO–VC	0.022	VC-Tower	0.000	LRAS3–Crew	0.030	VC–Crew	0.021	RVG-Tower	0.000	RVG–VC	0.017
Driver–Crew	0.022	VC-Driver	0.000	VC–Driver	0.024	RVG–Tower	0.016	RVG-Driver	0.000	RVO–Driver	0.017
LRAS3–RVG	0.019	RVG-RVO	0.000	Driver–VC	0.024	RVO–Driver	0.011	RVO-VC	0.000	Tower–Crew	0.017
VC–Driver	0.015	RVG-LRAS	0.000	LRAS3–VC	0.018	Tower–Crew	0.011	RVO-RVG	0.000	Driver–RVO	0.017
Driver–VC	0.015	RVG-Tower	0.000	LRAS3–RVG	0.018	Driver–RVO	0.011	RVO-LRAS	0.000	VC–LRAS3	0.008
Tower–Crew	0.011	RVG-Driver	0.000	Tower–Crew	0.018	VC–LRAS3	0.005	RVO-Crew	0.000	LRAS3–Crew	0.008
RVO–RVG	0.007	RVO-LRAS	0.000	RVO–Driver	0.012	Tower–RVO	0.005	RVO-Tower	0.000	Tower–RVO	0.008
RVO–Driver	0.007	RVO-Crew	0.000	Driver–RVO	0.012	Driver–Crew	0.005	RVO-Driver	0.000	Driver–Crew	0.008
Driver–RVO	0.007	RVO-Tower	0.000	Driver–LRAS3	0.012	RVG-RVO	0.000	LRAS-VC	0.000	RVG-RVO	0.000
Driver–LRAS3	0.007	RVO-Driver	0.000	RVG–LRAS3	0.006	RVG-LRAS	0.000	LRAS-RVG	0.000	RVG-LRAS	0.000
RVG–LRAS3	0.004	LRAS-RVO	0.000	RVG–Tower	0.006	RVG-Driver	0.000	LRAS-RVO	0.000	RVG-Driver	0.000
RVG–Tower	0.004	LRAS-Tower	0.000	RVO–RVG	0.006	RVO-RVG	0.000	LRAS-Tower	0.000	RVO-RVG	0.000
LRAS3–Driver	0.004	LRAS-Driver	0.000	LRAS3–Driver	0.006	RVO-LRAS	0.000	LRAS-Driver	0.000	RVO-LRAS	0.000
RVG-RVO	0.000	Tower-Crew	0.000	RVG-RVO	0.000	RVO-Tower	0.000	Tower-Crew	0.000	RVO-Tower	0.000
RVG-Driver	0.000	Tower-VC	0.000	RVG-Driver	0.000	LRAS-VC	0.000	Tower-VC	0.000	LRAS-VC	0.000
RVO-LRAS	0.000	Tower-RVG	0.000	RVO-LRAS	0.000	LRAS-RVG	0.000	Tower-RVG	0.000	LRAS-RVG	0.000
RVO-Tower	0.000	Tower-RVO	0.000	RVO-Tower	0.000	LRAS-RVO	0.000	Tower-RVO	0.000	LRAS-RVO	0.000
LRAS-RVO	0.000	Tower-LRAS	0.000	LRAS-RVO	0.000	LRAS-Tower	0.000	Tower-LRAS	0.000	LRAS-Tower	0.000
LRAS-Tower	0.000	Tower-Driver	0.000	LRAS-Tower	0.000	LRAS-Driver	0.000	Tower-Driver	0.000	LRAS-Driver	0.000
Tower-RVG	0.000	Driver-VC	0.000	Tower-RVG	0.000	Tower-RVG	0.000	Driver-VC	0.000	Tower-RVG	0.000
Tower-RVO	0.000	Driver-RVG	0.000	Tower-RVO	0.000	Tower-LRAS	0.000	Driver-RVG	0.000	Tower-LRAS	0.000
Tower-LRAS	0.000	Driver-RVO	0.000	Tower-LRAS	0.000	Tower-Driver	0.000	Driver-RVO	0.000	Tower-Driver	0.000
Tower-Driver	0.000	Driver-LRAS	0.000	Tower-Driver	0.000	Driver-RVG	0.000	Driver-LRAS	0.000	Driver-RVG	0.000
Driver-RVG	0.000	Driver-Crew	0.000	Driver-RVG	0.000	Driver-LRAS	0.000	Driver-Crew	0.000	Driver-LRAS	0.000
Driver-Tower	0.000	Driver-Tower	0.000	Driver-Tower	0.000	Driver-Tower	0.000	Driver-Tower	0.000	Driver-Tower	0.000

Finally, in these tables, if a given sender–receiver pair did not occur for a given condition, it is grayed out. In other words, if the RVG never spoke to the LRAS operator for a given condition, that pair is grayed out. From this, we can see in Table 5 (live run) the Army crew exhibited nine total sender–receiver pairs during target engagements, whereas the Marine crew exhibited 11 total sender–receiver pairs, although we note that those counts include sender–receiver pairs involving the Tower, which provided oversight for the gunnery task and is not considered a member of the gunnery crew.

Formatting data in the manner depicted in these tables provides several immediate insights into the crews' coordination patterns. First, the distribution of sender–receiver pairs that transmitted any information is narrower during engagements than during interengagement time. This means crew communication patterns were significantly restricted during the actual targeting and firing task so that task-directed information was communicated as efficiently as possible and distractions minimized. These patterns relaxed during interengagement time, which allowed for a "debrief" period in which crew members could interact more casually to review prior performance, discuss how to interact with the user displays, plan the next encounter, or even make a joke.

Second, Tables 4 and 5 provide a clear visual depiction of the crew partnerships. For example, in this data set it is apparent the VC and RVG have the most bidirectional communication patterns during the actual engagements and account for a large portion of the information flowing within the crew as a whole. This may have implications for the team's trust and cohesion. Trust is developed across repeated interactions between two parties (Rempel et al. 1985), so higher rates of bidirectional interaction in a given sender–receiver pair can be considered a marker of trust development over time. Thus, it can be posited that over enough team interactions, two team members that frequently form sender–receiver pairs would develop trust more quickly than they would with other teammates. This may also have implications for the trust and cohesion of the entire team. For example, the interactions of the VC and RVG appear to be most important to the team's performance. If communication between those two crew members goes well, the crew can perform well, resulting in better team states. However, breakdowns in the communication of the VC and RVG could result in knock-on effects on the crew's performance and, therefore, their perceptions of cohesion among the team.

Third, it is possible to identify structural differences between the Army and Marine information-flow patterns. For example, the Army crew communication patterns appeared to be more constrained to fewer sender–receiver pairs, with VC–RVG events accounting for 41.6% of all communication during live-fire engagements, compared with 32.6% for the Marine crew (Table 5). The disparity in communication patterns between the Army and Marine crews during gunnery engagements is even more pronounced in the dry-fire condition (Table 4), in which just five sender–receiver pairs in the Army crew accounted for all communication during target engagements, compared with 11 sender–receiver pairs for the Marine crew. Review of the transcripts suggests the Army crew's communication during the dry-fire runs was highly procedural with few divergences from their usual order of communicating gunnery commands. For example, during dry-fire engagements no crew member ever spoke directly to the LRAS operator, and the LRAS operator only ever provided information to the crew in general (e.g., "enemy troops in the open"). In contrast, during dry fire, the Marine crew exhibited more general information sharing (target locations, range readings, etc.) among the RVG, VC, and LRAS operator during the task.

These structural differences between crew-communication patterns may provide insight into team cohesion. Hung and Gatica-Perez (2010) found that highly cohesive teams in small group meetings 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, though it remains to be seen whether the findings of Hung and Gatica-Perez (2010) can apply to teams in other contexts, such as human–autonomy teams in military contexts. Based on our data, the Marine crew communication was more balanced during live fire in the sense there were more permutations of sender–receiver pairs in the gunnery crew: 10 sender–receiver pairs were noted during target engagements (the 11th pair, Tower–Crew, is excluded from this example given that the Tower is not a member of the gunnery crew). In contrast, the Army crew had fewer sender–receiver pairs, with seven sender–receiver pairs (this number likewise excludes the pairs Tower–Crew and VC–Tower in Table 5). The differences in the two crews' performance is most likely related to the increased gunnery experience held by the Army crew members, who were previously qualified in mounted gunnery operations. In addition, their communication patterns appear to reflect that experience more so than any difference in team cohesion, given that both crews were familiar prior to the Wingman JCTD. However, the ability to investigate communication patterns and aspects such as conversational dominance/balance may be ultimately useful to assessing the cohesion of human–autonomy teams with further development of this approach.

3.2 Relational Event Models

To compare communication dynamics across all four cases—Soldier (dry), Soldier (live), Marine (dry), and Marine (live)—four identical relational event models were run on the transcribed data. By comparing results across the type of crew and type of run, it is possible to identify similarities and differences in communication network structure for each condition (although we do not assess message content with this approach). In these relational event models, four families of related model terms were used (i.e., each family captures similar communication dynamics). These families capture effects related to the number of communication channels one has, one's tendency to continue utilizing commonly used communication channels, salience of communication channels, and conversational timing of interactions.

The first family of model terms captures preferential attachment, defined as the tendency for individuals with many communication channels to continue being involved in a disproportionately large amount of communication (Barabási and Albert 1999; Newman 2001). This theory suggests those with many communication partners may send or receive messages at higher rates, as their popularity may drive subsequent communication. Preferential attachment model terms are used here to identify whether message-sending is more concentrated or dispersed in the team. This helps provide insight into whether the group is using more centralized coordination.

The second family of model terms captures routinization, or the tendency for individuals to continue using communication channels they have used in the past. For example, as individuals repeatedly engage in behaviors in group contexts, routines related to those specific behaviors or communication patterns are often observed (Stinchcombe 1965; Wilson et al. 2007). To the extent an individual consistently communicates with a particular partner, one may continue to observe communication to that partner as that particular communication channel builds inertia over time.

The third family of model terms assesses recency effects, which suggest that more recently utilized communication channels become more salient. For example, after sending a message to a partner or receiving a message from that partner, an individual's next action (regardless of what else transpires among other actors in the network during the interim) may be to send a message to that partner because that remains the most salient interaction. While the three families of model terms discussed so far capture features of the network, the following family of terms focuses on specific sequences of interaction.

The final family of model terms captures participation shifts, or time-ordered sequences of communication (Gibson 2003). As individuals engage in conversation and follow conversational norms, they repeatedly shift into and out of three roles: speaker, recipient, and unaddressed third party. Gibson's notational scheme illustrates sequences of pairs of interactions that may occur among up to four individuals: A, B, X, and Y. (While the number of potential senders or receivers may greatly exceed four, a pair of dyadic interactions necessarily involves no more than four individuals: two senders and two receivers.) Gibson's notation indicates the sender–receiver pair of the first interaction (always A–B) followed by the sender–receiver pair of the second interaction. AB–BA illustrates, for example, that A directed an interaction to B, followed by B's reciprocal interaction with A. For this task, five other participation shifts are identified. AB–BY represents a potential handoff of information from A to B and then from B to Y. AB–AY represents coordinated disbursement of information as A communicates to B and then A communicates to Y. The final three participation-shift terms fall into Gibson's category of "turn usurping", in which a new speaker interrupts the call and response pattern of interaction. For example, AB–XA represents an interaction from A to B, followed by a third party's (X) interaction with A. This may occur if X needs to communicate with A based on what A said to B (perhaps inquiring how A obtained the information transmitted to B). AB–XB represents an A–B interaction followed by X's interaction with B. In this instance, X may have additional, complementary information to provide to B based on what A transmitted to B. The final turn-claiming interaction is AB–XY, in which two conversations may be occurring in parallel. By capturing these sequences of interaction, participation-shift terms provide useful insight into how information tends to flow throughout a group. In conjunction with the other terms, the relational event model provides insight into the timing and structure of communication in a network.

3.2.1 Results

In Table 6 the results from each of the four models are presented with the model coefficients grouped into families of related model terms. Coefficients should be interpreted similarly to those from survival analysis, where the coefficient represents the hazard, or likelihood, of an event of interest occurring per unit of time. In this case, the event of interest is a communication act. The coefficients therefore represent the rate of occurrence for specific types of communication. A coefficient with a positive, significant effect indicates an increased hazard of that particular type of communication act occurring while negative effects indicate a reduced hazard of that particular communication act. In the last row of Table 6, for example, applying the exponential function to each coefficient shows that an AB–XY (i.e., consecutive interactions between two separate dyads) has a relative hazard of 7.17 ($e^{1.97}$) in the Marines' dry scenario (compared with a non-AB–XY communication event), 0.075 in the Marines' live-fire scenario, 20.491 in the Soldiers' dry scenario, and 0.139 in the Soldiers' live-fire scenario. These results suggest a greatly increased rate of such events during the dry scenarios and a greatly reduced rate of such events during the live-fire scenarios. The magnitude of effect size differs across the Marine and Soldier teams, but the directions of the effects are identical. These types of comparisons across scenario types (dry vs. live fire) and across team (Marine vs. Soldier) help to demonstrate whether differences may arise due to differences in team or due to differences in the scenario type.

Table 6 Relational event model's results

Model term family	Term	Marine–dry coef. (SE)	Marine–live coef. (SE)	Soldier–dry coef. (SE)	Soldier–live coef. (SE)
Preferential attachment	Indegree–S	–0.56 (1.55)	–2.35 (0.86)[b]	6.28 (1.60)[c]	–1.17 (0.58)[a]
	Indegree–R	–0.83 (1.32)	–1.02 (0.90)	1.32 (1.95)	–2.61 (0.79)[c]
	Outdegree–S	7.13 (1.16)[c]	0.13 (0.97)	6.80 (1.60)[c]	1.09 (0.79)
	Outdegree–R	1.90 (1.41)	–7.75 (1.28)[c]	–5.23 (3.00)	–7.16 (1.10)[c]
Routinization	Inertia: S–S	3.47 (0.86)[c]	1.24 (0.54)[a]	4.28 (1.56)[b]	2.98 (0.68)[c]
	Inertia: R–S	–0.95 (0.87)	1.91 (0.77)[a]	6.04 (2.67)[a]	5.91 (0.81)[c]
Salience	Recency: R–S	–0.49 (0.73)	–0.51 (0.54)	–2.43 (2.23)	–3.19 (0.57)[c]
	Recency: S–S	0.05 (0.62)	2.28 (0.40)[c]	–0.49 (0.90)	0.60 (0.50)
Turn-receiving	PS: AB–BA	3.92 (0.64)[c]	1.83 (0.34)[c]	3.01 (0.65)[c]	1.82 (0.34)[c]
	PS: AB–BY	1.93 (0.72)[b]	–0.91 (0.40)[a]	3.71 (0.63)[c]	–0.52 (0.32)
Turn-continuing	PS: AB–AY	1.43 (0.67)[a]	–0.49 (0.38)	–0.02 (0.81)	–2.39 (0.46)[c]
Turn-usurping	PS: AB–XA	2.59 (0.61)[c]	1.66 (0.27)[c]	–0.50 (0.73)	0.01 (0.27)
	PS: AB–XB	2.52 (0.67)[c]	–0.48 (0.33)	1.75 (0.80)[a]	–1.15 (0.25)[c]
	PS: AB–XY	1.97 (0.69)[b]	–2.59 (0.39)[c]	3.02 (0.65)[c]	–1.97 (0.29)[c]

coef. = coefficient; SE = standard error; S = send; R = receive; PS = participation shift

[a] $p < 0.05$

[b] $p < 0.01$

[c] $p < 0.001$

3.2.2 Preferential Attachment Findings

In the first four rows of Table 6, the preferential attachment terms generally show consistency across teams but reveal differences across the dry- and live-fire scenarios. The table organizes model terms by incoming communication channels (indegree) or outgoing communication channels (outdegree) and by one's propensity to send (S) or receive (R) messages. Both dry scenarios show strong, positive, significant effects for the propensity for those with more outbound communication channels to send messages at higher rates. That is, those who have broadcast messages to several others will continue broadcasting at much higher rates than chance. The Soldier team shows a tendency for individuals with larger volumes of *incoming* communication channels to be more likely to send messages. These effects suggest a tendency for the teams to use centralized coordination in which a highly connected individual tends to relay information throughout the team. Across both teams we observe a notable change in the live-fire scenario.

Whereas the dry scenarios showed increased message-sending from those with more communication channels, the live-fire scenarios tended to show a *decrease* in activity from those with more communication channels. During the live-fire exercise, all of the significant coefficients in the set of preferential attachment terms point in the negative direction. In both teams individuals with more inbound communication channels send messages at lower rates and those with more outbound communication channels receive messages at lower rates. That is, individuals with communication channels that allow them to broadcast information do not receive many messages, and those who tend to receive more information do not broadcast information. These findings suggest a decentralization of information transmission within the team, as those with fewer channels become more likely to send messages. Together with the dry scenario results, these findings suggest both teams adopted more centralized coordination during dry runs and more decentralized coordination during live-fire exercises. These findings complement the communication-flow results to provide insight into how particular sequences of interaction and network-based patterns of interaction changed between dry- and live-fire scenarios. The differences between communication dynamics in the dry scenario and in the live-fire scenario were consistent for both the Marine team and Soldier team despite the performance differences between them. This suggests these changes in communication reflect changes in the changing environment rather than changes in performance outcome. Longer term, it will be important to map out how those patterns and differences are exhibited in high- and low-performing teams across different mission events, contexts, and milestones.

3.2.3 Routinization Findings

The next set of results comes from the two rows in the "Routinization" section of Table 6. In all models we find inertial effects for sending messages; if *i* sends many messages to *j*, then *i* is likely to continue sending messages to *j*. However, the strength of this effect is lower in the live-fire scenario than it is in the dry scenario. Except for the Marines' dry scenario, results show consistent reciprocation of inertial messages. That is, if *i* sends many messages to *j*, then *j* is more likely to send a message back to *i*. These tied inertial findings suggest routinized usage of communication channels for both teams during dry- and live-fire exercises. While the direction of the effect is the same for both teams, the routinization effects are consistently stronger for the Soldier team than for the Marine team. While we have insufficient statistical power to determine if this accounts for differences in team performance, routinization may be a useful predictor for investigating performance outcomes.

3.2.4 Salience Effects

The contrast between the crews' salience effects provides insight into how they differ in their handling of corrections for missed targets during the live-fire scenario. During the live-fire scenario the Marine crew had a tendency for *i* to send multiple, consecutive messages to *j*, as indicated by the positive, significant send–send recency coefficient. More specifically, the VC typically gave multiple instructions to help the RVG locate and aim for a target. The sample transcript in Fig. 2 demonstrates an example in which VC says "Yup," then "To the right," then "Your other right" as RVG prepares to hit a target. In cases where RVG misses the target, VC uses this same strategy. For example, after a miss VC says "still short" and follows that with "see if you can change to [a distance of] 800 [meters]". These sorts of sequential corrections were not present in the Soldier crew-communication logs. Instead, they had a negative receive–send recency effect, indicating that if *j* most recently received a message from *i*, then *j* has a reduced hazard of sending a message back to *i*. For example, when the RVG missed a target, the VC or the LRAS3 operator (often, both) would often provide some correcting information such as "still high" or "come down another half target from". The RVG typically did not respond to either correction but subsequently announces "on the way" to the whole crew to confirm the next shot is about to be fired. Because the dry-fire runs feature no live firing, we do not observe such corrections in those scenarios. These differences in communication strategies, particularly in response to misses, may provide insight into the performance differences between the Army and Marine teams. With a larger population of teams we would be better able to distinguish performance-enhancing communication patterns from those that may be artifacts stemming from differences in how Army and Marine teams are trained to communicate.

3.2.5 Participation-Shift Effects

Like the salience effects, the participation shift effects showcased notable differences between the dry- and live-fire scenarios, as well as some distinctions between the two teams. Beginning with the turn-receiving participation shifts, strong, positive effects for AB–BA (reciprocation) and AB–BY (relaying information) during both teams' dry run followed by a decline in effect size during the live-fire run are found. The AB–BA coefficient remains positive and significant for both teams, but AB–BY becomes nonsignificant for the Soldiers and negative for the Marines. A decline in effect size for the turn-continuing participation shift (AB–AY) from dry runs to live-fire runs is also seen. Although the reduction in effect size was similar across both teams, the magnitude of the effect differed and may account for their performance differences.

Like the turn-receiving and turn-continuing coefficients, the turn-usurping coefficients tend to be lower in live-fire runs compared with dry runs. We observed negative effects for AB–XY interactions during live-fire runs for both crews. This was in contrast to their positive effects during the dry-fire runs. This indicates a propensity against having multiple conversations occurring simultaneously. Examples of this AB–XY interaction are seen in both the Army and Marine crews during the dry-fire runs when the LRAS3 operator announces to the crew that a target is lased, followed by the VC providing some additional information (usually range) directly to the RVG. During the live-fire exercise, the Marine crew deviated from this pattern by following a "lased" announcement (from LRAS3 operator to crew) with a confirmation (from VC to LRAS3 operator). This demonstrates an example of the positive AB–XA turn-usurping pattern present in the Marine team only. By contrast, in the Soldier team the LRAS3 operator primarily confirmed hits or provided guidance for misses, followed by AB–BA interactions between VC and RVG to align correctly for the next attempt. In fact, the reciprocal AB–BA effect was the only positive effect for the Army crew and reflects a fairly basic pattern of call–response interactions during the live-fire scenario, in contrast to the Marine team's propensity to complement AB–BA with turn-usurping AB–XA interactions. The latter reflects a more structured pattern of communication than the Soldier team's higher levels of redundancy and duplication of messages. The Soldier team's better performance suggests these differences in patterns of participation shifts are worth examining as determinants of team performance.

3.2.6 Results Summary

In sum, the relational event model highlights distinctions between the Soldier and Marine crews as well as the dry- and live-fire runs. Both used more centralized coordination during the dry-fire runs and more decentralized coordination during live-fire exercises. Additionally, both teams demonstrated more routinized use of communication channels during the live-fire run. However, during the live-fire runs we observed patterns suggesting more redundancy of messages and more role overlap (primarily related to messages correcting for missed targets) by the Army crew while the Marine crew exhibited clearer roles with less redundancy. The introduction of live-fire played an important role in identifying distinctions between the teams, as much of this differentiation emerged when teams corrected for missed targets. Such distinctions in team-communication strategies may reflect differences in team trust and cohesion. Links between communication dynamics and measures of trust and cohesion deserve further attention in future research.

In practice, these findings suggest that observed communication patterns during dry-fire runs will not necessarily generalize to the live-fire exercises. Without capturing how teams respond to missed shots (which did not occur during dry runs), it is not possible to observe their strategies for adapting to mistakes. That adaptation ended up being a primary source of differentiation between the two teams. Additional training of the crews to ensure they are making the most of their dry-fire runs should include subsequent fire commands to incorporate activities they are likely to see during live-fire events. Vehicle commanders should not just announce "Target, cease fire" after the RVG has engaged the targets. Instead, a subsequent fire command will have the RVG reengage the target and practice what to do in case the rounds miss their target on the first burst. The tower could play a larger role during the dry run and randomly insert a sensing to the crew of over, short, left or right, which will cause the crew to react to a missed target during an engagement.

3.3 Linguistic Similarity Findings

3.3.1 Were Metrics of Linguistic Similarity Similar to Each Other?

Although LSM, LSS, and CASSIM were developed to address similar research questions, their methods are distinct enough to warrant concern that they reflect different constructs. Therefore, whether they correlated with each other and whether their associations differed if measured at the level of the entire crew (Table 7) and at the level of dyadic communication (i.e., for all dialogue between any two people; Table 8) were examined.

Table 7 Correlations of the linguistic-similarity measures at crew level

	LSM	**LSS**
LSS	0.370[a]	…
CASSIM	0.132[a]	0.182[a]

[a] a$p < 0.001$

Table 8 Correlations of linguistic-similarity measures at dyad level

	LSM	**LSS**
LSS	0.905[a]	. . .
CASSIM	–0.049	0.101

[a] a$p < 0.001$

At the crew level, LSM and LSS were positively related, r = 0.37, but not nearly at the level of these measures at the dyad level. At the dyad level, LSM and LSS were nearly identical, r = 0.905, whereas CASSIM was related to neither LSM nor LSS.

CASSIM appeared to be more consistently related to LSM and LSS when measured at the crew level, but their relationships were still rather small, r = 0.132 and r = 0.182, respectively. These patterns suggest LSM, LSS, and CASSIM may capture unique constructs regarding interpersonal communication.

3.3.2 How Synchronous Were Dyads within the Crews?

As was made apparent in the analysis of communication networks, some dyads within a larger crew communicate regularly whereas others rarely, if ever, interact. These results show whether certain dyads exhibited greater levels of synchrony and whether these assessments were dependent on the method of measurement. Findings in Table 9 reveal inconsistencies in the available information for each dyad across teams and runs, as well as inconsistencies in how synchronous dyads are across teams and runs. Of the dyads for which there was enough information to evaluate synchrony at each level, communications between RVG and VC were the only roles exhibiting consistent synchrony regardless of crew, run, or metric used to assess synchrony. The other roles were either unequally represented across teams and runs or inconsistent in their levels of synchrony.

Table 9 Linguistic measures of each dyad, in each run, within each crew

Team	Run	Dyad	LSM	LSS	CASSIM
Soldier	Dry fire	Driver–RVO	0.001	NA	0.727
		Driver–VC	0.588	0.883	0.602
		RVG–VC	0.359	0.379	0.652
		RVO–VC	0.479	0.735	0.662
	Live fire	Driver–RVO	0.567	0.715	NA
		Driver–VC	0.281	0.555	NA
		LRAS–RVG	0.156	0.375	0.656
		LRAS–VC	0.473	0.693	0.735
		RVG–VC	0.523	0.805	0.632
		RVO–VC	0.004	NA	0.750
Marine	Dry fire	Driver–LRAS3	0.022	0.195	0.659
		Driver–RVO	0.001	NA	1
		Driver–VC	0.688	0.766	0.613
		LRAS–RVG	0.095	0.548	0.643
		LRAS–VC	0.527	0.812	0.720
		RVG–VC	0.793	0.965	0.704
		RVO–VC	0.471	0.752	0.696
	Live fire	Driver–LRAS3	0.001	NA	0.692
		Driver–RVG	0.548	0.957	0.895
		Driver–RVO	0.576	0.867	0.601
		LRAS–RVG	0.600	0.702	0.633
		LRAS–RVO	0.115	0.535	0.569
		LRAS–VC	0.624	0.866	0.666
		RVG–VC	0.783	0.963	0.672
		RVO–VC	0.164	0.465	0.692

Note: Values correspond to dyadic synchrony in communication for each team and each run. LSM measures the proportion of function words for all communication between the specified dyad; LSS measures the cosine similarity in discussed topics, as extracted by LSA; and CASSIM measures the inverse of the edit distance between sentence structures.

When reading the communications between RVG and VC, we find their communications are frequent but they often contain the same types of communication that follows a format of VC telling RVG the location of targets, then RVG reporting they identified and fired on the targets. Within the Army crew, the VC tended to give precise directions (e.g., “Gunner troops, 12 o’clock, 500 meters”), to which the RVG responded succinctly (e.g., “Roger”). For the Marine crew, the VC and RVG tended to engage in more casual conversation, even though their conversation was still related to the task. Take, for instance, the following exchange between Marines during the dry-fire run:

> VC: RVG, how's that look on your screen with the stuff I have populated for you? Is that too much? Does it even help?
>
> RVG: It kinda does because I know what area I'm at.
>
> VC: 'Cause you don't know what area you're in?
>
> RVG: Yeah.

In this example, the Marine VC and RVG were trying to understand how to share information effectively. Exchanges like these contribute to higher levels of linguistic synchrony, but they may be irrelevant, or possibly even detrimental, to team performance in the short term. Considering that Soldiers outperformed Marines in the gunnery task, the Soldiers' lower levels of linguistic synchrony may indicate this measure is not fit for assessing how well the team will perform on the task. However, interactions like the one between the Marine VC and RVG are crucial for building common knowledge regarding their tasks and capabilities, which in turn creates the conditions for effective trust and cohesion. To understand the relevance of dyadic linguistic synchrony, multiple teams with similar knowledge states about the task would need to be examined, which will deconfound the effects of linguistic synchrony from the effects of the type of team.

3.3.3 How Synchronous Were Teams?

Linguistic synchrony for each team, run, and engagement was assessed. At a general level, results show whether teams differed in synchrony across live- and dry-fire events. No outstanding levels of synchrony were observed, although the Army crew during live fire appeared more synchronous according to LSM and LSS, but not so much according to CASSIM (see Table 10).

Table 10 Synchrony for each crew and run, averaged across engagements

Team	Engagement	LSM	LSS	CASSIM
Soldier	Dry fire	0.6894	0.3814	0.5205
	Live fire	0.8658	0.4163	0.6512
Marine	Dry fire	0.6638	0.3622	0.6585
	Live fire	0.6603	0.3325	0.6889

Note: Values correspond to team-level synchrony in communication for each run. LSM measures the proportion of function words for all communication among members of a team during the specified run; LSS measures the cosine similarity in discussed topics, as extracted by LSA; and CASSIM measures the inverse of the edit distance between sentence structures.

Although the Army team shows some distinctions between dry and live fire according to LSM and CASSIM, the Marines team shows similar synchrony across both exercises. Whereas the Marines showed a distinct change in the network structure of their communications according to the relational event models, the synchrony of their communication was little changed. The combination of these findings demonstrates that the network structure of communication will not necessarily predict features of the content (e.g., synchrony) and vice versa. As such, a comprehensive analyses of intrateam communication likely requires both analyses.

To examine crew-level synchrony across engagements, a plot that displays the trajectories of linguistic synchrony over the course of engagements was generated, separately for each team and each run (Fig. 4).

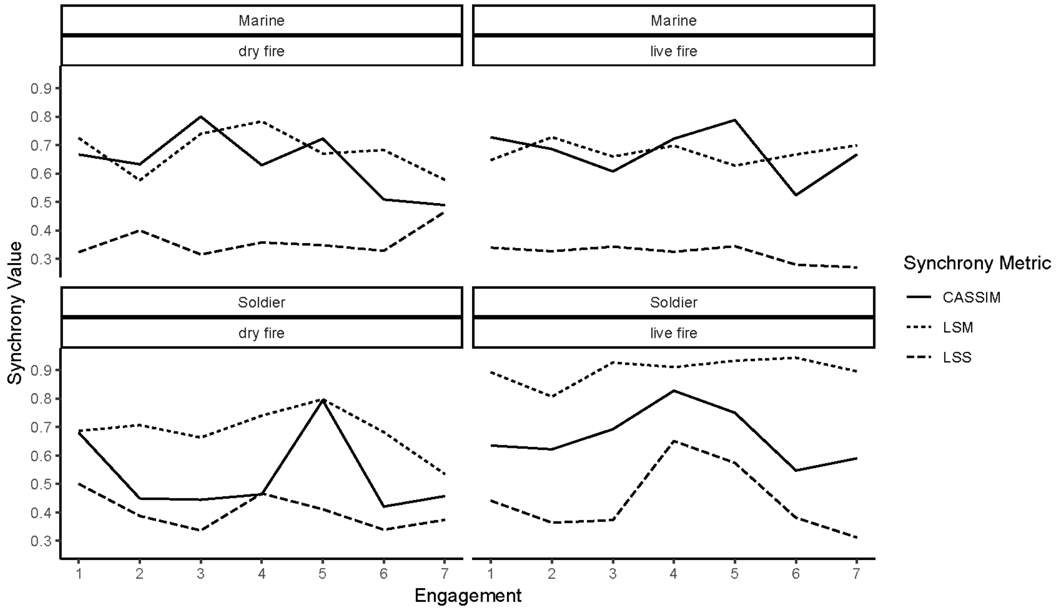


Fig. 4 Trajectories of linguistic synchrony across engagements, separated by teams (Marine and Soldier) and exercises (live fire and dry fire)

For a given team one might expect linguistic synchrony to increase over time, as people are becoming more familiar with their roles; however, as can be seen in Fig. 4, there was no synchrony metric for which a clear linear pattern emerged, nor did any pattern appear similar across scenarios or teams. Further, despite the strong correlations between LSM and LSS, their absolute values were very distinct across these levels of comparison, with LSM values being consistently higher than LSS. Further, despite the relatively low correlation between LSM and CASSIM, their magnitudes appeared quite comparable, which creates further difficulties in understanding how these metrics relate to each other.

4. Implications for Team-Trust and Team-Cohesion Metrics

Team communication provides a unique window into the functioning of a team. The approaches described in this report lay the groundwork to derive team trust and cohesion insights from the ways crews naturally communicate, which can then inform our understanding of a team's performance. Team trust and cohesion are developed over many repeated interactions among crew members. Because aggregate communication flow depicts the proportion of team communication accounted for by each pair of crew members, it may be possible to determine the development of trust and cohesion by following the strength of communication among crew members. For example, if aggregate communication flow reveals that one crew member is relatively isolated during on-task and off-task phases, it stands to reason they may experience a lower feeling of cohesion than a crew member who serves a more centralized role. Research will need to investigate the extent to which these communication links can relate to perceptions of trust and cohesion in both novel and intact human–autonomy teams. Further, it will be important to understand how crew-communication patterns relate to their performance in a given scenario or context. This will shed additional light on how their communication can affect trust and cohesion, given that some aspects of team trust and cohesion are affected by team outcomes.

In addition to time-aggregated communication as a means for characterizing team member perceptions of trust and cohesion, disaggregated communications modeled with relational event models may provide insight into those team states. Team states tend to arise and evolve as a function of communication (Marks et al. 2001), and certain types of interactions—centralized coordination, routinized communication, salient interactions, or certain types of participation shifts—may shape trust and cohesion. Capturing communication dynamics over time, relational event models may be instrumental in inferring those states as a function of interaction dynamics. By linking changes in those states to features of communication expressed via a relational event model's coefficients, future research may be able to identify these links between team states and communication network dynamics.

By focusing on whether people are communicating similarly, whether it be similarity of content (e.g., words or topics) or similarity of syntax, we ought to be able to understand the extent to which team members are effectively building common knowledge. For team members to work effectively, especially when their roles are interdependent, they must build a common understanding of how the information transmitted among members relates to their environments (Thomas et al. 2014). The methods employed here use various techniques to evaluate whether people are using words that are themselves indicative of common knowledge

(LSM), whether people are explicitly discussing similar topics (LSS), and whether people are adopting similar sentence structures (CASSIM). Each of these methods contributes a distinct metric for understanding which aspects of communication teammates are converging.

Overall, more work is needed in this area to make these communication metrics usable for human–autonomy teams, especially in a field setting. One research direction will involve developing an understanding of how synchrony and asynchrony relate to cohesion in various contexts. In some team contexts, it may be the case that a high degree of asynchrony could relate to effective teamwork, such as in cases where team members have a high degree of independence with little overlap in tasking and a reduced reliance on tight coordination. Therefore, more research will build our understanding of how synchrony (or asynchrony) can relate to performance, trust, and cohesion as well as the team conditions that affect the relationship between those factors.

Additionally, more work is needed on technological advances to provide the capabilities for real-time data acquisition and analysis methods. For the linguistic-similarity approach to be implemented in real-time, we would need 1) an accurate speech-to-text tool that reliably extracts relevant speech in diverse environments; 2) the ability to identify who is talking to whom (e.g., speech channels to control or detect whether speaking to individuals, subgroups, or the whole team), which is essential for understanding dyadic synchrony; and 3) an understanding of how to update these metrics in real time.

For real-time analysis, it is currently unclear how much data are needed and how often the text should be analyzed, as these analyses are conducted based on the proportion of words that fall into a special dictionary (LSM), how words cluster together (LSS), or the structure of sentences (CASSIM). As the document (i.e., transcribed speech) develops over the course of a conversation, so too will the proportion of certain types of words, the clustering of words, and the structure of sentences. Due to the nature of these computations, they may not be informative of real-time performance but rather informative of team dynamics over the span of long, stable relationships among teammates. As trust and cohesion are themselves emergent properties of interactions over time, these metrics of linguistic synchrony are poised to yield important insights regarding how teams build common knowledge.

In addition to resolving the technical aspects of real-time data acquisition and analysis, the current data set is limited by a lack of information. In other words, there were relatively few communication events among several team members, so it is difficult to know whether their contributions to team-level synchrony were informative of overall team dynamics. The sparseness of these data makes it impossible to test whether the variability we see is systematic or mere noise. To overcome this drawback, we aim to collect data over longer timespans and from more teams. Although we collected additional metrics that scholars argue should be related to linguistic synchrony (e.g., team performance and self-reported trust), drawing relationships between linguistic synchrony and those other metrics is untenable without substantially more data.

In future contexts, other aspects of communication such as speech emotion data may be used to identify times when autonomous aids are needed. Assuming that team cohesion within teams composed of human and autonomous entities is similar to that of human teams, team cohesion would improve in teams with access to such adaptive autonomous assistance.

5. Conclusion

The purpose of this report was to identify and assess possible communication-based metrics for assessing trust and cohesion in human–autonomy teams: a) communication flow assessment, b) relational event modeling, and c) linguistic similarity. Our data suggest it is possible to derive trust and cohesion from teams' communication content. However, these approaches were limited by their reliance on transcribed and coded text data. Manual transcription and coding is highly time-intensive, and in order to be useful for future HAT scenarios, automated methods that work in near real-time will be necessary to support those methods.

Communication-flow assessment provided a useful window into the aggregate communication behaviors of a crew. Clear differences were found between the sender–receiver pairs exhibited during engagements versus during inter-engagement time, noting that crews significantly restricted their communication structures while on-task. Although the method used in this report involved assessing communication events (i.e., logging senders and receivers) manually, approaches for quickly deriving communication events are being developed to support future analyses that rely on sender–receiver data that can feed automated analyses of communication patterns (Baker et al. 2019). This method also provides a possible window into team trust and cohesion by revealing the extent to which different pairs of crew members communicate with each other. Because trust and team cohesion are developed across repeated interactions over time, aggregate communication counts should map to the development of those states over time, though more work is needed to understand how a given amount of communication can result in a change in trust or cohesion.

Clear distinctions between how teams communicate in dry- and live-fire training scenarios are shown through relational event models. The changes in model results reflect how teams issued corrections following misses during live-fire scenarios, a situation they did not encounter during dry scenarios. Additionally, differences in the magnitude of effect sizes across teams may explain some of their performance differences. More performance outcomes coupled with communication logs will give us the statistical power to identify relationships between communication dynamics and performance outcomes. Further, relational event models may be useful to infer team states such as trust or cohesion as a function of interaction dynamics. Follow-up work will develop our understanding of how changes in trust and cohesion map to changes in features of communication expressed via a relational event model's coefficients.

Noticeable trends regarding how different team members, and different teams as a whole, tend to communicate were identified through linguistic similarity. Although there were apparent divergences between the synchrony metrics, indicating they are capturing unique features of communication, there were notable convergences, as well. Namely, communications between the VC and RVG appeared to be marked by consistently higher levels of synchrony, at least compared with other dyads on the team, which may suggest the synchrony of a subset of members could explain disproportionate variability in trust, cohesion, and performance at the team level. The current data set had a relatively small number of communication events from which to draw linguistic insights; therefore, in future testing we will collect data over longer timespans and from more teams to improve our ability to detect linguistic synchrony effects.

In all, the primary takeaway from these analyses is that the results are promising for using communication metrics on human–autonomy crew interactions to gain an understanding of team performance, trust, and cohesion. Our methods identified differing patterns of communication across teams and across dry- and live-fire scenarios. These patterns provided insight into how crews coordinated and performed their tasks, which subsequently provided a lens through which their trust and cohesion could be explored. However, more work is needed to understand exactly how these metrics relate to trust and performance in human–autonomy teams. The advantage of the current data set that fed our analyses was that it was derived from a live human–autonomy teaming exercise in the field; however, the approaches would have benefited from more communication events to better understand how these metrics trend with performance, trust, and cohesion. Given this, our current work is preliminary in nature, and we expect to extend these approaches to upcoming field and simulation studies to refine their ability to produce important, meaningful insights into the trust, cohesion, and performance of human–autonomy teams in various contexts.

6. References

[ARCIC] US Army Capabilities Integration Center. Robotics and autonomous systems strategy. Fort Eustis (VA): Army Capabilities Integration Center (US); 2017 Mar [accessed 2019 Aug 22]. http://www.arcic.army.mil/App_Documents/RAS_Strategy.pdf.

Babcock MJ, Ta VP, Ickes W. Latent semantic similarity and language style matching in initial dyadic interactions. J Lang Soc Psych. 2014;33(1):78–88.

Baker AL, Schaefer KE, Hill SG. Teamwork and communication methods and metrics for human–autonomy teaming. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); 2019 Oct. Report No.: ARL-TR-8844.

Barabási AL, Albert R. 1999. Emergence of scaling in random networks. Science. 286(5439):509–512.

Barnes M, Chen JYC, Schaefer KE, Kelley T, Giammanco C, Hill SG. Five requisites for human-agent decision sharing in military environments. In: Savage-Knepshield P, Chen J, editors. Advances in human factors in robots and unmanned systems. Basel (Switzerland): Springer International Publishing; 2017. p. 39–48.

Boghrati R, Hoover J, Johnson KM, Garten J, Dehghani M. Conversation level syntax similarity metric. Behav Res Methods. 2018;50(3):1055–1073.

Butts CT. 2008. Social network analysis: a methodological introduction. Asian J Soc Psych. 11(1):13–41.

Cannon-Bowers JA, Salas E, Converse S. Shared mental models in expert team decision making. In: Castellan NJ, editor. Individual and group decision making: current issues. Hillsdale (NJ): Psychology Press; 1993. p. 221–246.

Clark HH, Brennan SE. Grounding in communication. In: Resnick LB, Levine J, Teasley SD, editors. Perspectives on socially shared cognition. Washington (DC): American Psychological Association; 1991. p. 127–149.

DeCostanza AH, Marathe AR, Bohannon A, Evans AW, Palazzolo ET, Metcalfe JS, McDowell K. Enhancing human–agent teaming with individualized, adaptive technologies: a discussion of critical scientific questions. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2018 May. Report No.: ARL-TR-8359.

Demir M, McNeese NJ, Cooke NJ. Team synchrony in human–autonomy teaming. In: Chen J, editor. Advances in human factors in robots and unmanned systems. Proceedings of International Conference on Applied Human Factors and Ergonomics; 2017 July 17–21; Los Angeles, CA. Springer, Cham. p. 303–312.

[DOA] Headquarters, Department of the Army. Training and qualification, crew. Washington (DC): Headquarters, Department of the Army (US); 2015 Mar. Training Circular No.: TC 3-20.31.

Fischer U, McDonnell L, Orasanu J. Linguistic correlates of team performance: toward a tool for monitoring team functioning during space missions. Aviat Space Env Med. 2007;78(5):B86–B95.

Gibson DR. Participation shifts: order and differentiation in group conversation. Soc Forces. 2003;81(4):1335–1380.

Gonzales AL, Hancock JT, Pennebaker JW. Language style matching as a predictor of social dynamics in small groups. Commun Res. 2010;37(1):3–19.

Hung H, Gatica-Perez D. Estimating cohesion in small groups using audio-visual nonverbal behavior. IEEE Trans Multi. 2010;12(6):563–575.

Janis IL. Victims of groupthink: a psychological study of foreign-policy decisions and fiascoes. Boston (MA): Houghton Mifflin; 1972.

Kiekel PA, Cooke NJ, Foltz PW, Shope SM. Automating measurement of team cognition through analysis of communication data. In: Smith MJ, Koubek RJ, Salvendy G, Harris D, editors. Usability evaluation and interface design: cognitive engineering, intelligent agents, and virtual reality. Mahwah (NJ): Erlbaum Associates; 2001. p. 1382–1386.

Klein D, Manning CD. Accurate unlexicalized parsing. ACL ’03: 41st Annual Meeting on Association for Computational Linguistics, vol. 1; Stroudsburg (PA): Association for Computational Linguistics; 2003. p. 423–430.

Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. Dis Proc. 1998;25(2–3):259–284.

Macrae C. Human factors at sea: common patterns of error in groundings and collisions. Marit Policy Manag. 2009;36(1):21–38.

Marathe AR, Schaefer KE, Evans AW, Metcalfe JS. Bidirectional communication for effective human-agent teaming. In: Chen JYC, Fragomeni G, editors. Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation. Springer International Publishing; 2018. p. 338–50.

Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team processes. Acad Man Rev. 2001;26(3):356–376.

Mullen B, Anthony T, Salas E, Driskell JE. Group cohesiveness and quality of decision making: an integration of tests of the groupthink hypothesis. Small Group Res. 1994;25(2):189–204.

Navarro G. A guided tour to approximate string matching. ACM Comp Sur. 2001;33(1):31–88.

Newman MEJ. Clustering and preferential attachment in growing networks. Phys Rev E. 2001;64(2):02510-2–025102-4.

Niederhoffer KG, Pennebaker JW. Linguistic style matching in social interaction. J Lang Soc Psych. 2002;21(4):337–360.

Rempel JK, Holmes JG, Zanna MP. Trust in close relationships. J Person Soc Psych. 1985;49(1):95.

Sacks H, Schegloff EA, Jefferson G. A simplest systematics for the organization of turn-taking for conversation. Language. 1974;50(4):696–735.

Salas E, Sims DE, Burke CS. Is there a “big five” in teamwork? Small Group Res. 2005;36(5):555–599.

Schaefer KE, Baker AL, Brewer RW, Patton D, Canady J, Metcalfe JS. Assessing multi-agent human-autonomy teams: US Army robotic Wingman gunnery operations. SPIE Defense + Commercial Sensing; 2019a; Baltimore, MD.

Schaefer KE, Brewer RW, Baker AL, Pursel ER, 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. Aberdeen Proving Ground (MD): CCDC Army Research Laboratory (US); 2019b. Report No.: ARL-TR-8663. https://apps.dtic.mil/docs/citations/AD1069401.

Sexton JB, Helmreich RL. Analyzing cockpit communication: the links between language, performance, error, and workload. Hum Perf Extrem Environ. 2000;5(1):63–68.

Snook SA. Friendly fire: the accidental shootdown of U.S. Black Hawks over northern Iraq. Princeton (NJ): Princeton University Press; 2000.

Stinchcombe AL. Social structure and organizations. In: March JG, editor. Handbook of organizations. Chicago (IL): Rand McNally; 1965. p. 142–193.

Thomas KA, DeScioli P, Haque OS, Pinker S. The psychology of coordination and common knowledge. J Pers Soc Psych. 2014;107(4):657–676.

Tiferes J, Hussein AA, Bisantz A, Kozlowski JD, Sharif MA, Winder NM, Ahmad N, Allers J, Cavuoto L, Guru KA. The loud surgeon behind the console: understanding team activities during robot-assisted surgery. J Surg Ed. 2016;73(3):504–512.

Wildman JL, Salas E, Scott CPR. Measuring cognition in teams: a cross-domain review. Hum Fact. 2014;56(5):911–941.

Wilson JM, Goodman PS, Cronin MA. Group learning. Acad Man Rev. 2007;32(4):1041–1059.

List of Symbols, Abbreviations, and Acronyms

ARL	Army Research Laboratory
C2	command and control
CASSIM	ConversAtion level Syntax SImilarity Metric
CCDC	US Army Combat Capabilities Development Command
CSTS	Combat Skills Training School
DIDEA	detect, identify, decide, engage, assess
ERP	Essential Research Program
HAT	human–autonomy teaming
HMMWV	High-Mobility Multipurpose Wheeled Vehicles
JCTD	Joint Capability Technology Demonstration
LRAS3	Long-Range Advanced Scout Surveillance System
LSA	latent semantic analysis
LSM	Linguistic Style Matching
LSS	Latent Semantic Similarity
NA	not available/not available
PS	participation shift
R	receive
RCV	robotic combat vehicle
RVG	robotic-vehicle gunner
RVO	robotic-vehicle operator
S	send
SE	standard error
TC	Training Circular
TET	target engagement trial
TTPs	tactics, techniques, and procedures
VC	vehicle commander

1 DEFENSE TECHNICAL
(PDF) INFORMATION CTR
DTIC OCA

1 CCDC ARL
(PDF) FCDD RLD CL
TECH LIB

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

1 CCDC ARL
(PDF) FCDD HSI
J THOMAS

1 USAF 711 HPW
(PDF) 711 HPW/RH
K GEISS

1 USN ONR
(PDF) ONR CODE 341
J TANGNEY

1 USA NSRDEC
(PDF) RDNS D
D TAMILIO

1 OSD OUSD ATL
(PDF) HPT&B
JB PETRO

<u>ABERDEEN PROVING GROUND</u>

18 CCDC ARL
(PDF) FCDD RLH
J LANE
Y CHEN
P FRANASZCZUK
K MCDOWELL
K OIE
FCDD RLH BD
D HEADLEY
FCDD RLH FA
A DECOSTANZA
S M FITZHUGH
D FORSTER
FCDD RLH FB
A EVANS
FCDD RLH FC
J GASTON
FCDD RLH FD
A MARATHE
A L BAKER
A KRAUSMAN
A SCHARINE
K E SCHAEFER
FCDD RLV A
R W BREWER
G SLIPHER

2 NSWC
(PDF) R L BEALE
E R PURSEL

1 USCENTCOM
(PDF) T LESTER

2 GROUND VEHICLES
(PDF) SYSTEMS CENTER
RTI GVR
T MICHALIK
K BRIGGS

1 ARMAMENTS CENTER
(PDF) J GIACCHI