Exploration of the Impact of Interpersonal Communication and Coordination Dynamics on Team Effectiveness in Human-Machine Teams

Article *in* International Journal of Human-Computer Interaction · November 2022

DOI: 10.1080/10447318.2022.2143004

Mustafa Demir
Arizona State University

Myke C. Cohen
Arizona State University

Craig Johnson
Arizona State University

Erin Chiou
Arizona State University

Taylor & Francis
Taylor & Francis Group

International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

Exploration of the Impact of Interpersonal Communication and Coordination Dynamics on Team Effectiveness in Human-Machine Teams

Mustafa Demir, Myke Cohen, Craig J. Johnson, Erin K. Chiou & Nancy J. Cooke

To cite this article: Mustafa Demir, Myke Cohen, Craig J. Johnson, Erin K. Chiou & Nancy J. Cooke (2022): Exploration of the Impact of Interpersonal Communication and Coordination Dynamics on Team Effectiveness in Human-Machine Teams, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2022.2143004

To link to this article: https://doi.org/10.1080/10447318.2022.2143004

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2022.2143004

Exploration of the Impact of Interpersonal Communication and Coordination Dynamics on Team Effectiveness in Human-Machine Teams

Mustafa Demir[a], Myke Cohen[b], Craig J. Johnson[b], Erin K. Chiou[b], and Nancy J. Cooke[a,b]

[a]Center for Human, AI, and Robot Teaming, Global Security Initiative, Arizona State University, Tempe, AZ, USA; [b]Human Systems Engineering, Arizona State University, Mesa, AZ, USA

ABSTRACT
Teams composed of human and machine members operating in complex task environments must effectively interact in response to information flow while adapting to environmental changes. This study investigates how interpersonal coordination dynamics between team members are associated with team performance and shared situation awareness in a simulated urban search and rescue (USAR) task. More specifically, this study investigates (1) how communication recurrence affected and reflected coordination dynamics between a USAR robot and human operator when they used different communication strategies, and (2) how these dynamic characteristics of the human–robot interpersonal coordination were associated with the team performance and shared situation awareness. The USAR interpersonal coordination dynamics were systematically characterized using discrete recurrence quantification analysis. Results from this study indicate that (1) teams demonstrating more flexibility in their coordination dynamics were more adaptive to changes in the task environment, and (2) while robot explanations help to improve shared situation awareness, revisiting the same communication pattern (i.e., routine coordination) was associated with better team performance, but did not improve shared situation awareness.

1. Introduction

With advancements in data mining and machine learning algorithms, machines (i.e., Artificial Intelligence -AI, synthetic teammates, automated agents, and robots) are poised to revolutionize work in almost every industry. Various examples indicate that AI-enabled machines are becoming nearly universally important to command and control missions. For instance, integrated wearable electronics with large-scale machine control can allow warfighters access to innumerable possible contingencies in critical situations, such as the "F-35 Joint Strike Fighter"—a fighter plane and flying sensor in one—which condenses vast amounts of data from the environment into visualizations in the pilot's helmet (Pellerin, 2015). High-impact applications can also be found in healthcare and space exploration. For example, to protect healthcare workers from the coronavirus pandemic (COVID-19), robots have been deployed to help care for patients at hospitals in Italy (Romero, 2020) and team with health providers to deliver telehealth services in China (Hornyak, 2020). The extent to which machines can aid human endeavors over great geospatial and temporal distances is perhaps best exemplified by the Perseverance rover and other robots distributed on Mars (mars.nasa.gov, n.d.).

In the context of increasingly capable machines, recent work has used a team science lens to gain insight into how machines can be designed to succeed in more dynamic task domains. This team science lens acknowledges the strengths of teams, the emergent qualities of teaming, and the support for interdependencies needed for effective teaming. Therefore, the team science lens effectively considers how machines functioning more like active team members rather than passive tools would need to be designed. The current study focuses on human-machine teams (HMTs) in the dynamic task context of urban search and rescue (USAR). Previous work has demonstrated that in this context, communication and coordination within HMTs is vital for effective task and teamwork (Chiou et al., 2022; Demir et al., 2017; Demir, Likens, et al., 2019; Demir, Amazeen, et al., 2020; Demir, McNeese, et al., 2020; Johnson et al., 2021; Kopp et al., 2022; Nalepka et al., 2021; Shah & Breazeal, 2010).

In this study, we first define HMTs and explain interpersonal team communication and coordination in the context of nonlinear dynamical systems (NDS). Then, we systematically analyze how effective teamwork can be achieved in terms of communication (specifically robot explanations) and coordination within HMTs. Finally, we discuss how our findings can guide the design of future HMTs.

1.1. Human-machine teaming and teamwork

A team can be defined as two or more entities occupying heterogeneous roles that work interdependently toward a common goal or task (Cannon-Bowers et al., 1995); HMTs

CONTACT Mustafa Demir mdemir@asu.edu Human Systems Engineering, Arizona State University, 150A Santa Catalina, Mesa, AZ 85212, USA

satisfy this definition while being comprised of humans and interactive machines. HMTs are different from more traditional configurations of humans and technology (e.g., supervisory control) in that machine team members need to communicate and coordinate with human teammates according to task interdependencies and may also be placed in team structures in which machines have to "team" with other machines on top of communicating and coordinating with humans (O'Neill et al., 2022).

Before discussing the intricacies of HMTs as a process, it is important to define two components of teaming as a process: taskwork and teamwork. Taskwork refers to individual responsibilities that do not require the participation of other teammates, but must be executed well for the team's goals to be realized (Salas et al., 2008). On the other hand, teamwork refers to interdependent behaviors, such as social processes, enacted by multiple team members to achieve common goals—it is essentially referring to the shared, goal-oriented interactions among teammates (Marks et al., 2001). Accordingly, a team's success depends on its team member's skills, which contribute to the team processes involved in accomplishing teamwork (Marks et al., 2001). Therefore, a team member would need to be able to perform stable taskwork to contribute to teamwork, meeting with the needs of other team members in the process. Consider, for instance, a simple restaurant consisting of a server and a cook forming a dyadic team. The server's taskwork includes interacting with customers and taking their orders, while the cook's individual-level task is to prepare food for the server to bring to diners. Teamwork between the two refers to their collective ability to achieve their shared goals, such as consistently serving all customers high quality meals in a timely manner. Their teamwork can be observed in the staff's willingness to negotiate the timing of new orders, including giving estimates about delayed orders if any, and relaying customer feedback about served meals. In other words, teamwork is often best represented by team members' interactions with one another (Cooke et al., 2013).

Although teamwork has been framed both as behavior and as a collective cognitive process (Salas et al., 1995), it is generally agreed upon that an essential process of teamwork is team coordination (Brannick et al., 1995). As such, a teammate whose singular role is to facilitate teamwork, such as a communication control center operator, has to make reliable decisions at an individual level (e.g., who should receive a certain message, which messages must be filtered out or modified) to ensure effective interactions between teammates.

It is not unusual for taskwork and teamwork components in complex team task environments to be at odds with each other, particularly when it comes to the allocation of limited resources (Chiou & Lee, 2016). The task of balancing the two is a complex process. In all-human teams, this can be mitigated by harnessing social intelligence and capability to predict the needs, beliefs, and intentions of other teammates—abilities that machines are yet to have (Stowers et al., 2021). Issues in prioritizing taskwork over teamwork and vice versa are known to cause subpar HMT performance compared to when the same machines and humans are placed in purely machine or purely human teams, respectively (OpenAI et al., 2019). This may partly be due to machine shortcomings when interacting with human teammates, which affect the latter's cooperative behaviors towards it and attitudes like trust (Chiou & Lee, 2016, 2021). Such limitations in machine-teaming capabilities may cause disruptions in established taskwork, and teamwork interdependencies may even lead to breakdowns in human-machine coordination beyond repair, resulting in catastrophic disasters. An example of this is the Patriot Missile incident where British and American confederates were shot down in fratricide by the U.S. Army (Cummings, 2006).

One way to study coordination dynamics—including breakdowns therein—is through looking at communication data (e.g., instant messaging, transcripts from audio recordings), which has been used as real-time, non-obtrusive measures of team performance in the past (Gorman et al., 2012). A benefit of using real-time data is that information loss during post-processing of communication is minimized. For example, nonlinear dynamical systems (NDS) analysis of team communication data has been used to assess team situation awareness (TSA) in real-time (Grimm et al., 2018). Recent work has also shown that it may be possible to use real-time communication data as indirect online measures of attitudes towards machine teammates, such as changes in personifying and objectifying language indicating changes in trust levels (Cohen et al., 2021). It is implicit, then, that team performance as measured by the accomplishment of teamwork goals may also be predicted in real-time by looking at communication between teammates, which could be the basis for novel interventions to aid maladaptive teams. Rooted in this, the following section discusses interpersonal communication and coordination.

1.2. Interpersonal communication and explanations

Interpersonal communication is an essential element of team coordination (Cooke et al., 2013); it entails both the quality of communicated content and "the sequencing of messages in conversations and the sequencing of conversations into relationships" (Pearce, 1976, p. 17). Additionally, interpersonal communication that is of analytical interest within collaborative settings (a) is intentional, (b) is influenced by psychological, technological, and environmental factors, and (c) occurs only within certain relational contexts, i.e., there should be a common goal to achieve (Gudykunst, 2000). Communicators apply their knowledge of actions to construct output representations that reflect their plans for accomplishing a team task. These representations are structured and consist of goals, ideas to be communicated, suitable language, and neural commands required to produce the language (Cappella, 1987). One key element of interpersonal communication that is structured as such is the *explanation*.

Explanations are explicit communications that provide a contrastive reason behind a decision or action occurring relative to a teammate's understanding and are often instantiated as responses to "why" questions (Miller, 2019). Explanations should be clear, accurate, and explicit to reduce task uncertainty, and should clarify the issue or behaviors in question among the team members. Otherwise, misunderstandings between teammates can arise, task uncertainty and problems may continue, and in turn, adversely affect situation awareness and performance (de Visser et al., 2020). Explanations can also support building shared mental models with robot teammates, preventing unexpected challenges, and improving timely adaptation to team task environments (Miller, 2021). A previous study analyzing data from the same experiment as this study discovered that a moderate level of robot explanations contributed to better shared situation awareness and team performance in a dynamic task environment (Chiou et al., 2022). In the current study, we focus on how varying levels of the quality of robot explanations, can lead to different team coordination patterns.

1.3. Interpersonal coordination and nonlinear dynamical systems

What team members say and how they say it seems to help them develop a narrative for successful coordination (Arthur et al., 2020), and in turn, effective teamwork (Demir, Likens, et al., 2019). Coordination is defined as dependency management and can occur in biological, social, and physical systems (Malone & Crowston, 1994). It can involve managing the components and processes of a system that change over time (Butner et al., 2014), and can be observed on different levels and across substrates (Butler, 2011), such as team communication (Louwerse et al., 2012) and physiology (Likens et al., 2014; Palumbo et al., 2017). Relatedly, *interpersonal coordination* refers to how behavior, physiology, emotional, and cognitive states covary between individuals over time (Wiltshire et al., 2020). Researchers have linked communicative aspects of interpersonal coordination to important constructs and outcomes in teamwork, such as flow of the communication (Chanel et al., 2013; Demir, Likens, et al., 2019; Gorman et al., 2016; Guastello & Peressini, 2017; Wiltshire et al., 2018). Interpersonal coordination is an essential measure of the dynamical systems properties of teamwork, both in all-human teams (Gorman et al., 2010; Guastello, 2010; Guastello & Guastello, 1998; Likens et al., 2014; Ramos-Villagrasa et al., 2018; Schmidt & Richardson, 2008; Wiltshire et al., 2020) and in HMTs (Demir, Cooke, et al., 2018; Demir, McNeese, et al., 2018b; Fiore & Wiltshire, 2016; Nalepka et al., 2021).

Teams can be described as continuously evolving dynamical systems, such that team coordination has been characterized using NDS methods (Gorman et al., 2010; Guastello, 2010). With NDS methods, it is also possible to characterize team coordination dynamics between team members and their task environment by using coordination of communication as an input (Demir, McNeese, et al., 2018a). In general, an NDS is a system that continuously evolves through its behavior, and its emergent behavior is the result of its system components interacting over time (Thelen & Smith, 2007). In teams, specifically, many behavioral patterns within the system can emerge, and transitions between patterns are often sudden and nonlinear (Perone & Simmering, 2017). An NDS can behave in many ways, but all possible behaviors fall within a multidimensional "state space." Suppose the system develops some repeated behavior patterns. In that case, that means it favors a region of the state space and is said to have moved into an "attractor state," which a system is likely to return to sometime after departing it (Spencer & Perone, 2008). Teamwork behaviors and patterns can emerge as teammates coordinate with one another, as each member sets out to fulfill their respective roles while also synchronizing with other team members to achieve their common goal. This leads to emergent team behavior that is ideally robust, flexible, and fault-tolerant over time and under routine or novel conditions (Maes, 1993).

Many studies have examined all-human team coordination using NDS methods to identify how the dynamical properties of team coordination relate to factors such as team composition and various kinds of task perturbations (Gorman et al., 2012; Gorman et al., 2010; Gorman et al., 2012; Schmidt & Richardson, 2008; Wiltshire et al., 2018). A commonly used NDS method in these studies is Recurrence Quantification Analysis (RQA). RQA quantifies the number and length of recurrent patternspresent using a state-space trajectory in a dynamical system (Fusaroli & Tylén, 2016; Gorman et al., 2012; Russell et al., 2012; Strang et al., 2014; Marwan et al., 2007). Researchers have previously investigated team coordination in the context of HMTs by using RQA and its multivariate extensions (Demir, McNeese, et al., 2018a, 2019, 2020; Gorman et al., 2020; Nalepka et al., 2021). For instance, Demir, Likens, et al. (2019) showed that HMTs should be "flexible" to adapt to dynamic task uncertainty and also be "stable" to maintain performance (for more elaboration regarding stability and flexibility, see Demir, Likens, et al., 2019). Similar studies have also used RQA to observe how calibrated coordination (i.e., adjusting the coordination based on the task situation) relates to better team performance (Demir, Cooke, et al., 2018; Demir, McNeese, et al., 2019, 2020) and team situation awareness (Nalepka et al., 2021). However, previous uses of RQA in the literature have focused broadly on coordination between the team members, but in this study, we specifically manipulate robot explanations to see how interpersonal coordination changes and relates to team effectiveness.

This study examines the association between HMT interpersonal coordination dynamics, robot explanations, and team effectiveness (i.e., performance and shared situation awareness) in a USAR simulation. Therefore, there are two research questions. First, we addressed how communication recurrence affects and reflects interpersonal coordination dynamics between the USAR robot and human operator when using different levels of explanation. The second research question we addressed is how these dynamic characteristics of the human-robot teams were associated with team effectiveness metrics (i.e., shared situation awareness and team performance).

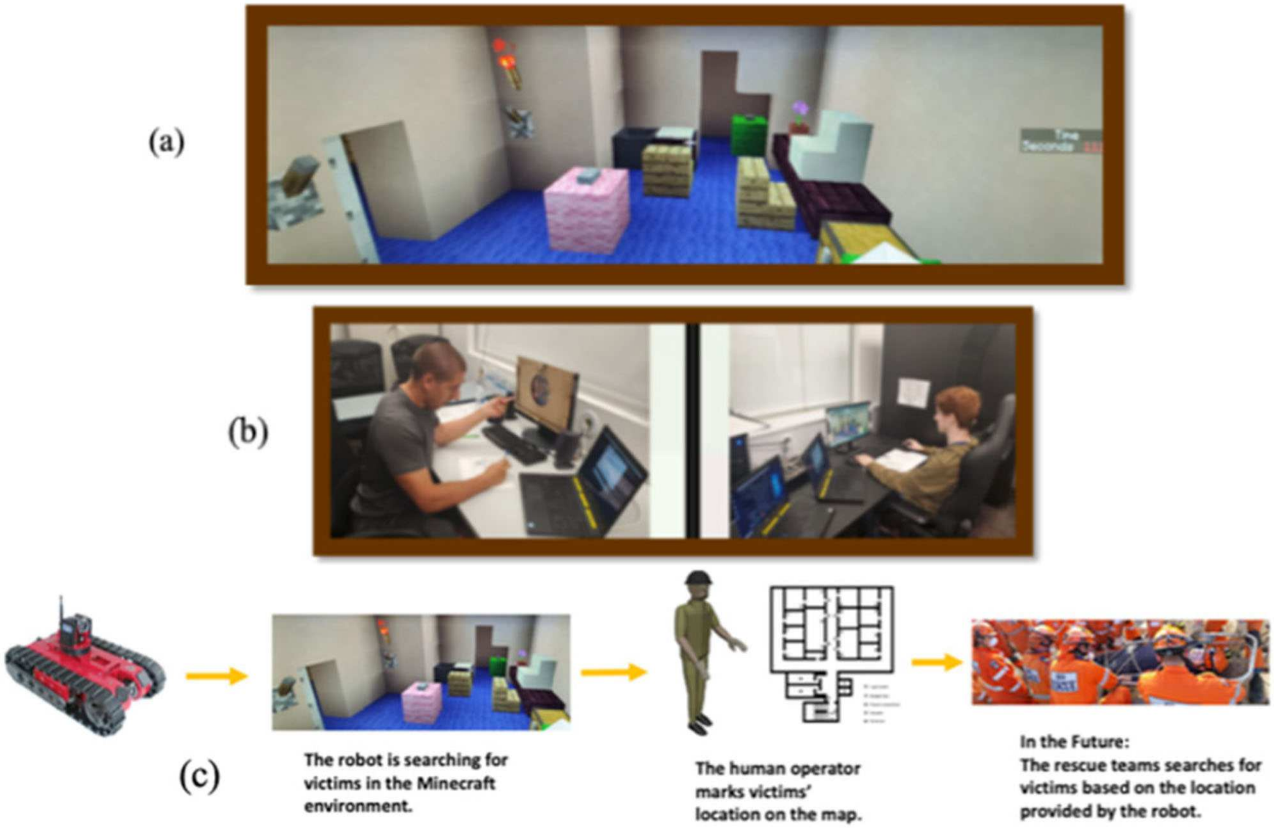


Figure 1. (a) A screen view of simulated Minecraft task environment, (b) team members' location (the randomly assigned participant on the left and the WoZ robot on the right), and (c) task sequence.

2. Task environment and experimental design

2.1. Simulated minecraft task environment

A simulated USAR task environment (Gonzalez et al., 2005) developed in Minecraft (version 1.11.2) and based on a previous microworld design (Bartlett & Cooke, 2015) was used to collect team communication data. This testbed afforded dynamic reconfigurations to the task environment (Lematta et al., 2019), requiring teams to adapt to the environment through text chat communication. The USAR task involved two team members: a "robot" team member that autonomously searched a partially collapsed building and located potential victims and a "navigator" who remotely monitored and documented key information through the robot's live video feed. The "robot" was, in fact, a highly-trained researcher controlling a first-person perspective virtual avatar in Minecraft. A check-list used by the other researcher to train the researcher who mimic the "robot," and the "robot" used script to send the messages in a timely manner (Chiou et al., 2022; Demir et al., 2017). This makes use of the "Wizard of Oz" (WoZ) paradigm popular in user-centered research (Kelley, 1983). Study participants were assigned to the navigator role and were told to complete two USAR missions with the autonomous robot. The navigator was to document the location of victims on a pre-collapsed floor map of the building and was also tasked with documenting any changes to the map (Figure 1). The goal was to complete a reconnaissance mission before sending in another team to recover victims safely.

In addition to the shared video feed, the robot interacted with the navigator via a text chat communication system. Participants were not limited to what they could ask or say to the robot through the text chat, but the robot was limited to only providing scripted explanations that were relevant to the task environment, and to respond to the participant using a series of pre-scripted phrases. For example, in response to questions that were not relevant to the task environment, the robot would reply with, "Sorry, I am unable to help you with that." The robot was purposefully designed not to have complete natural language abilities, so the scripted responses ensured some control across study conditions and believability that participants were indeed interacting with a robotic entity.

2.2. Experimental design

This study investigates how interpersonal coordination dynamics were related to team effectiveness metrics through a secondary analysis of data collected for a study on robot explanation strategies (Chiou et al., 2022). Four between-subject conditions tested three robot explanations strategies and one communication priming condition. In the *Always Explain* condition, the robot explained all plan deviations, and other explanations were provided when requested. In the *Explain If Asked* condition, all explanations were provided *only* when requested. In *Pull Prime*, participants received some training that prompted them to initiate asking the robot questions; otherwise, this was the same as the

Explain If Asked condition. In *Never Explain*, the robot only acknowledged, but never fulfilled requests for explanations.

3. Methods

3.1. Participants

An a priori power analysis was conducted using G*Power3 (Faul et al., 2007) to test the difference between the means of 4-conditions by 2-missions using an *F-test*, with a medium effect size ($\eta_p^2 = 0.06$; Cohen, 1977), and an alpha ($\propto$) of 0.05. According to the result, a total sample of 60 participants with four equal-sized groups of $n = 15$ was required to achieve a power of 0.90. A total of 60 participants ($M_{age} = 22.48$, $SD_{age} = 7.21$, $Male = 42$, $Female = 18$) from Arizona State University and the surrounding campus community participated in a 1.5-h session study (each participant was compensated $15). All participants were fluent in English, reported normal or corrected-to-normal hearing and vision (e.g., were not color blind), and reported having experience using a computer mouse and keyboard.

3.2. Procedure

Participants were tasked with completing a series of simulated reconnaissance missions as part of a human-machine team. The team's goals were to identify and triage trapped victims in a collapsed virtual building and annotate any structural changes to the building resulting from the collapse. After obtaining informed consent, each participant engaged in a 1 h and 10-min session that included a training session followed by two actual missions with a 20-min break in between the two missions. The training session included a 10-min voiced-over slide presentation that explained the task, the robot's role and capabilities, team goals and individual goals, how to communicate with the robot, and how performance would be scored. Following the slide presentation, participants took part in an additional 10-min training mission in which they observed the robot navigating a training version of the Minecraft environment (i.e., a simpler building plan) while practicing how to communicate with the robot.

After completing the training, participants completed two 20-min missions in which the participant workload increased from Mission 1 to Mission 2 (i.e., the number of critical victims in the same building increased to add temporal urgency and difficulty). Before each mission, participants received an instruction sheet reminding them of the mission goals, and a printed map of the pre-collapsed structure. The printed map also included information about the robot's planned search route, indicated by yellow arrows. During each mission, the participants used the map by marking victim locations and building changes on them as the robot navigated the environment and relayed its location to the participant (in real-time through its video feed and via text chat). After each mission, participants were asked to fill out the NASA Task Load Index (NASA TLX; Hart & Staveland, 1988) workload assessment and a questionnaire to assess trust in the robot. After the second mission, a separate questionnaire on participant demographics and perceptions about the missions was also administered. The robot's activity was obtained via game data through the Minecraft program and included for analysis.

3.3. Measures

The following team effectiveness (1–4) and team process (5–6) measures were obtained in this study: (1) shared situation awareness; (2) team performance; (3) trust in the robot based on the Jian et al. (2000) trust in automation scale (Jian et al., 2000; 4) perceived workload based on the NASA Task Load Index (Hart & Staveland, 1988; 5) communication behaviors that were coded as either information pushes (i.e., information sent by the robot) or pulls (i.e., information requested by the participant; Demir et al., 2017, and (6) communication flow. Demographics were also collected to assess the representativeness of the sample population. However, only the following measures were used for this particular study to explore how dyadic team interaction and robot explanations were related to team performance and shared situation awareness:

Shared situation awareness was measured through the participant's ability to accurately mark the location of survivors and building changes on the map in each mission. It is a percentage score that is calculated by summing the correctly annotated collapses and openings on the map, divided by the total number of collapses and openings, and multiplied by 100 (in terms of distribution, shared situation awareness is normally distributed). In a previous analysis of the same data set, the Always Explain condition teams were more accurate (had higher shared situation awareness) than the teams in the *Pull-Prime* condition but did not differ from the remaining two conditions (Chiou et al., 2022). A Shapiro-Wilk test (Shapiro & Wilk, 1965) showed a significant departure from normality, $W(120) = 0.94$, $p < 0.0001$. Accordingly, the data is negatively skewed (*skewness:* -0.49; see Figure 2). In order to meet the normality assumption, we applied an arcsine transformation (variance-stabilizing transformation; Lin & Xu, 2020) because of the data type (data are proportions ranging from 0 to 100). The transformation whose distribution is closer to normality (*skewness:* -0.03; see Figure 2), Shapiro-Wilk test $= W(120) = 0.95$, $p = 0.0003$. Therefore, later on, we also did regression diagnostics tests during the regression model-building steps (see the summary of the test results in Table 1).

Team performance was measured by calculating the proportion of correctly identified victims in terms of location and status. The number of victims triaged was counted, divided by the number of victims (total in the structure), and then multiplied by 100 for each mission. In a previous analysis of the same data, we reported that there was no significant difference across the robot communication conditions with respect to team performance ($p > 0.05$) (Chiou et al., 2022). In terms of distribution of the data, a Shapiro-Wilk test again showed a significant departure from normality, $W(120) = 0.92$, $p < 0.0001$, and a negative skew was found

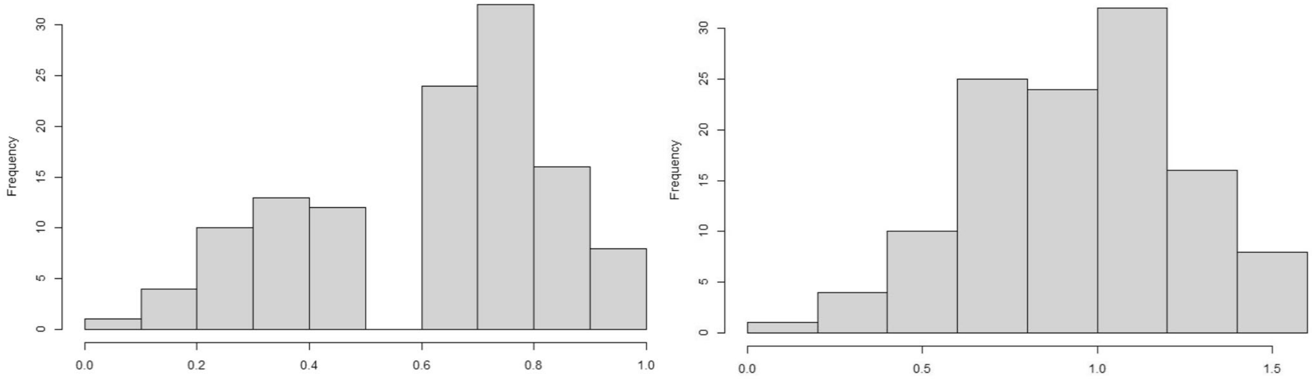
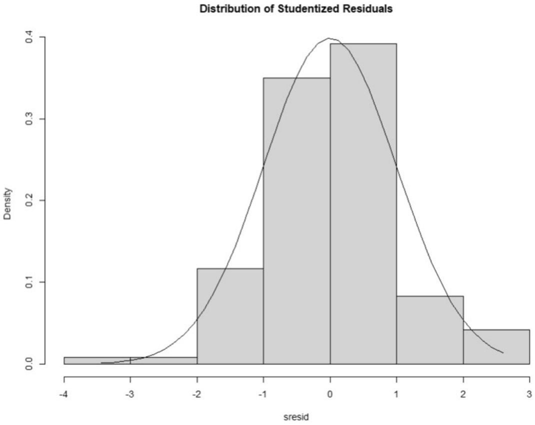


Figure 2. Shared situation awareness distribution: Left - before the transformation and right - after the transformation.

Table 1. Regression diagnostics.

Assessment	Description	Result
Outliers (Bonferroni Outlier Test) (*R package: CAR;* Fox et al., 2022)	Assesses a *t* distribution to test whether the model's largest studentized residual value's outlier status is statistically different from the other observations in the model. A significant *p*-value indicates an extreme outlier that warrants further examination.	$rstudent = -3.43$, $Bonferroni = 0.098$ (There was no extreme case to review, $p > 0.05$)
Normality (*R package: MASS;* Ripley, 2021)	Assesses how closely the model residuals resemble a normal distribution	 The histogram in the figure indicates that the residuals are approximately normally distributed.
Multicollinearity (*R package: MASS;* Ripley, 2021)	Assesses when two or more independent variables are highly correlated with one another in the model	*Recurrence Rate*$_{Vif}$ = 1.14 *Explanations*$_{Vif}$ = 1.14 None of the VIFs was bigger than 2. Therefore, there was no multicollinearity between Recurrence Rate and Explanations
Global Test of Model Assumptions (*R package: gvlma;* Pena, 2019)	Assesses the linear model assumptions, as well as performs specific directional tests designed to detect skewness, kurtosis, a nonlinear link function, and heteroscedasticity.	*Value* *p-value* *Decision* *Global Stat* 3.77 0.438 Acceptable *Skewness* 0.08 0.785 Acceptable *Kurtosis* 2.28 0.131 Acceptable *Link Function* 1.40 0.236 Acceptable *Heteroscedasticity* 0.01 0.893 Acceptable All the assumptions are acceptable.

for the data (*skewness:* −1.17; see Figure 3). We applied an arcsine transformation as in the shared situation awareness measure to arrive at distribution that is closer to normality (*skewness:* −0.14; see Figure 3 [right]), Shapiro-Wilk test $= W(120) = 0.95$, $p = 0.0003$. In Figure 3, it is noticeable that there are some extreme values for the transformed dataset; we tested and eliminated these by applying regression diagnostics while building the regression model (see Table 2).

Communication flow captures the flow of team members' communications. The instant message data were coded according to who sent a message. Navigator messages were coded as "1" and robot messages as "0". Across the conditions, the communication flow length slightly differs from one another: *Always Explain:* Mean = 51.1, $SD = 9.61$; *Explain If Asked*: Mean = 44.9, $SD = 9.60$; *Never Explain*: Mean = 46.1, $SD = 11.0$; *Pull Prime*: Mean = 58.4, $SD = 12.0$.

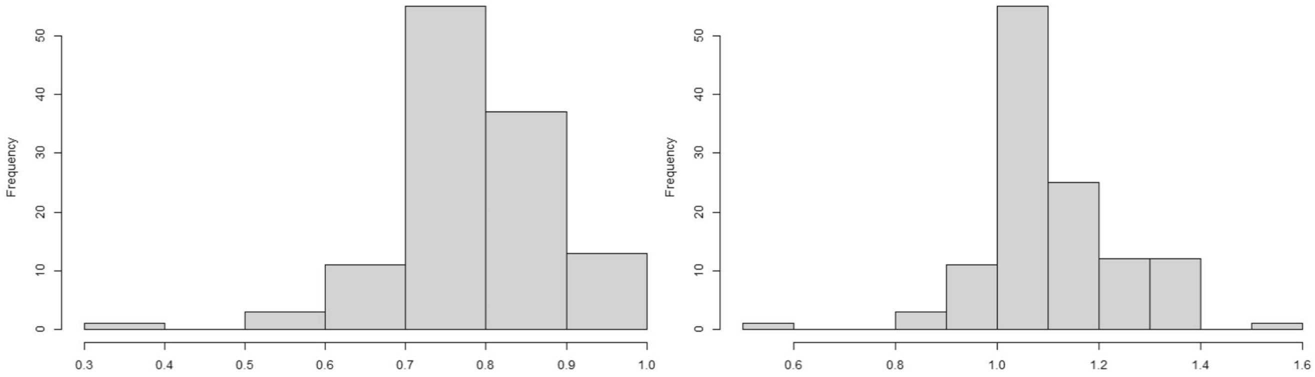
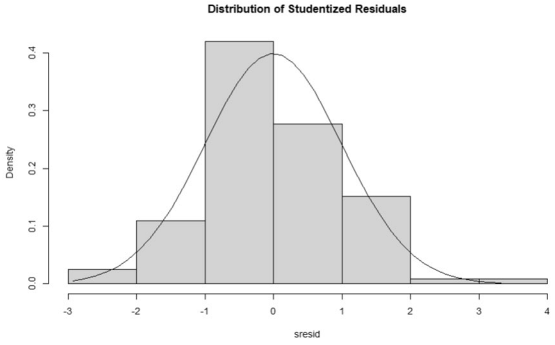


Figure 3. Team performance distribution: Left - before the transformation and right - after the transformation.

Table 2. Regression diagnostics.

<table>
<tr><th>Assessment</th><th>Description</th><th>Result</th></tr>
<tr><td><i>Outliers</i> (Bonferroni Outlier Test) (<i>R package: CAR;</i> Fox et al., 2022)</td><td>Assesses a <i>t</i> distribution to test whether the model's largest studentized residual value's outlier status is statistically different from the other observations in the model. A significant <i>p</i>-value indicates an extreme outlier that warrants further examination.</td><td>$rstudent = -4.06$, $Bonferroni = 0.011$
There were two extreme cases to review, $p < 0.05$.
We deleted one of the extreme cases on the dataset and re-run the regression and outlier analyses: $rstudent = 3.32$, $Bonferroni = 0.144$
No extreme case was detected, $p > 0.05$</td></tr>
<tr><td><i>Normality</i> (<i>R package: MASS;</i> Ripley, 2021)</td><td>Assesses how closely the model residuals resemble a normal distribution</td><td>

The histogram in the figure indicates that the residuals are approximately normally distributed.</td></tr>
<tr><td><i>Multicollinearity</i> (<i>R package: MASS;</i> Ripley, 2021)</td><td>Assesses when two or more independent variables are highly correlated with one another in the model</td><td><i>Recurrence Rate (Linear)</i>$_{Vif}$ = 1.55
<i>Recurrence Rate (Quadratic)</i>$_{Vif}$ = 1.52
<i>Maximum Length (Linear)</i>$_{Vif}$ = 1.56
<i>Determinism (Linear)</i>$_{Vif}$ = 1.43
None of the VIFs was bigger than 2. Therefore, there was no multicollinearity between the predictors in the model</td></tr>
<tr><td><i>Global Test of Model Assumptions</i> (<i>R package: gvlma;</i> Pena, 2019)</td><td>Assesses the linear model assumptions and performs specific directional tests designed to detect skewness, kurtosis, a nonlinear link function, and heteroscedasticity.</td><td><table>
<tr><th></th><th><i>Value</i></th><th><i>p-value</i></th><th><i>Decision</i></th></tr>
<tr><td><i>Global Stat</i></td><td>2.73</td><td>0.605</td><td>Acceptable</td></tr>
<tr><td><i>Skewness</i></td><td>0.70</td><td>0.403</td><td>Acceptable</td></tr>
<tr><td><i>Kurtosis</i></td><td>0.64</td><td>0.425</td><td>Acceptable</td></tr>
<tr><td><i>Link Function</i></td><td>1.07</td><td>0.304</td><td>Acceptable</td></tr>
<tr><td><i>Heteroscedasticity</i></td><td>0.33</td><td>0.564</td><td>Acceptable</td></tr>
</table>
All the assumptions are acceptable.</td></tr>
</table>

The number of explanations is coded as 1 when the robot provided explanations after the navigator asked a "why" question or an implied "why" question, such as "what about" questions or explanations which were directly given by the robot without waiting for the questions (i.e., *Always Explain* condition). In coding the explanations, there was almost perfect agreement between raters [$\kappa = 0.936$ (95% CI, 0.885–0.987)]. Therefore, we took the average of the two raters' codes and summed them across each mission.

4. Analytical perspective and results

4.1. Discrete recurrence quantification analysis

The team's interpersonal coordination dynamics were quantified using RQA on the communication flow data. Interactions were represented by a binary code that yielded a discrete, then used as an input into a discrete RQA. Discrete RQA quantifies dyadic team coordination processes and the dynamics that contribute to that process (Gorman

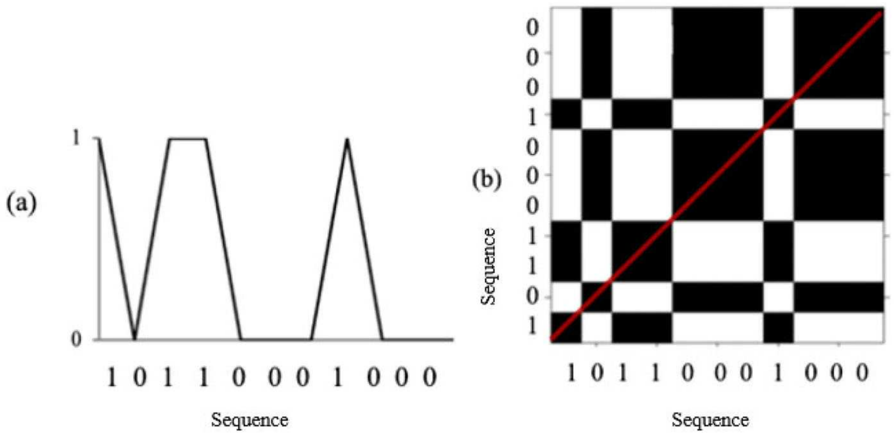


Figure 4. (**a**) Example Discrete Time Series, and (**b**) Discrete Recurrence Plot. Recurrent "points" (block boxes) are plotted whenever '1' repeats at a later time (reprinted from Demir et al., 2021).

et al., 2012), allowing for a measurement of the effect of perturbations (in this context, building changes) on a team's stability. Several measures were extracted from discrete RQA for this study, including percent determinism (*DET*), recurrence rate (*RR*), longest diagonal line (*MaxL*; i.e., stability), entropy (*ENTR*), laminarity, trapping time, and longest vertical line.

Figure 4(a) is a simple binary time series of length $N = 11$, $x(t) = [1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0]$ and Figure 4(b) is a visual representation, or recurrent points (RP) of the time series (Demir et al., 2021). Discrete RPs are constructed by placing a symbolic time series on both the horizontal and vertical axes of a graph and plotting a recurrent point (black box) whenever a symbol repeats. In this case, the value at $x(1)$ is repeated at $x(3)$, $x(4)$, and $x(8)$; likewise, the value at $x(3)$ is repeated at $x(4)$ and $x(8)$. The RP in Figure 2(b) gives a visual summary of these patterns, as well as repetitions involving zeros. The example concerning points, $x(1)$, $x(3)$, $x(4)$, and $x(8)$, are visually depicted in Figure 4(b) by tracing upwards from the bottom-left corner to the top-left corner of the plot, where "points" (black boxes) are plotted each time the value at $x(1)$ repeats at a later time in the series. The red line from the lower left-hand corner to the upper right-hand corner indicates the main diagonal. Only the upper triangle of the RP is analyzed because the matrix is symmetrical around the main diagonal. Recurrent points forming diagonals off the main diagonal indicate patterns that form when data segments match segments from earlier or later times (Marwan & Webber, 2015). This study used discrete RPs to examine the change in the following three commonly used RQA measures: *RR*, *DET*, *MaxL*, and *ENTR*. Note that these measures are content-independent measures that characterize the patterns of interaction (flow), rather than what is being talked about (Cooke & Gorman, 2009).

Recurrence Rate (*RR* or percent recurrences) measures the percentage of time a team revisits a communication pattern and captures the overall tendency for recurrence (i.e., recurrence density) on an RP. It is given by the ratio of the number of recurrent points to the square of the time series length. *RR* is calculated through the following formula (Marwan et al., 2007):

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,\ j=1}^{N} R_{i,\ j}(\varepsilon) \tag{1}$$

where R is the binary recurrence matrix of $\vec{x}$, N is length $(i,\ j = 1, \ldots,\ N,)$, and the similarity threshold is $\varepsilon \geq 0$ (Schultz et al., 2015). This formula can be interpreted as the probability of finding a recurrence trajectory $\vec{x}$ (reconstructed from a time series x, e.g., by time delay embedding, see Packard et al., 1980). *RR* quantifies the percentage of points that return to the same local neighborhood in the reconstructed phase space over time. An *RR* of 0% means the time series never revisit the same local neighborhood, whereas a rate of 100% means the time series revisit perfectly. In team coordination dynamics, we interpret percent recurrence as *a team's tendency to revisit a communication pattern* during teamwork.

Percent determinism (*DET*) is an index of how deterministic the structure of dyadic communication behavior is, calculated as the ratio of recurrence points forming diagonal lines to all recurrent points in the upper triangle of the RPs (Marwan et al., 2007). In this study, it was used in order to characterize how organized each team's communication behaviors were by measuring the distribution of recurrent points on the recurrence plot; highly organized systems are bound to repeat sequences of states many times, represented by many diagonal lines in an RP, whereas mildly deterministic systems would rarely do so and can be seen by the presence of only a few diagonal lines in an RP (Allen et al., 2017). DET was specifically calculated as follows (Marwan et al., 2007):

$$DET = \frac{\sum_{l=l_{\min}}^{N} lP(l)}{\sum_{l=1}^{N} lP(l)} \tag{2}$$

where "l" is the diagonal line length considered when its value is $\geq l_{min,}$ and $P(l)$ is the probability distribution of line lengths. A *DET* of 0% means the time series never repeats, whereas a rate of 100% means the time series repeats perfectly. We interpret DET as *a percentage of team coordination predictability* in this context.

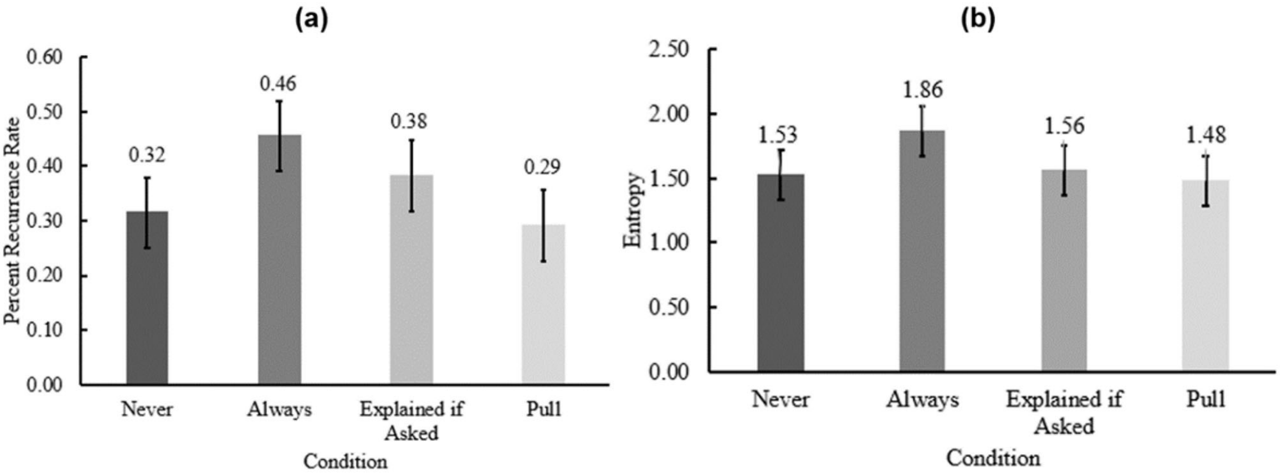


Figure 5. (a) Recurrence Rate and (b) entropy across the conditions (Error bars indicate 95% Confidence Interval).

Maximum Line Length (*MaxL*) is an indicator of the system's stability and captures the length of the longest diagonal sequence of recurrent states in the RP (Marwan et al., 2007). In an RP for a binary time series, it is simply the longest diagonal line. According to (Eckmann et al., 1987), the length of the diagonal lines is inversely proportional to the largest positive Lyapunov Exponent (i.e., index of attractor stability, see Abarbanel, 1996; Kantz & Schreiber, 1997; Stergiou, 2016). It is calculated as follows:

$$MaxL = \max\left(\{l_i\}_{i=1}^{N_l}\right) \tag{3}$$

where l_i is a diagonal line length which is a segment of the trajectory, and the total diagonal lines are $N_l = \sum_{l \geq l_{\min}} P(l)$ (Marwan et al., 2007). *MaxL* provides information about the *stability of team coordination*. Higher *MaxL* indicates higher stability and vice versa.

Entropy (*ENTR*) refers to the Shannon entropy of the probability $p(l) = P(l)/N_l$ to find a diagonal line of exactly length l in the RP:

$$ENTR = -\sum_{l=\ l_{\min}}^{N} p(l) \ln p(l) \tag{4}$$

ENTR reflects the complexity of the RP in respect of the diagonal lines. Lower entropy represents lower *complexity in team coordination* and vice versa.

4.2. Split plot analysis of variance

In order to address how conditions differ according to each of the team coordination dynamics measures across the missions, we applied split-plot Analysis of Variance (ANOVA). Recurrence rate findings show that there was a significant condition main effect, $F(3,\ 60)\ =\ 5.21$, $p = 0.003$, though there was no mission main effect, $F(1,\ 60)\ =\ 1.16$, $p = 0.692$, nor an interaction effect of condition by mission, $F(3,\ 60)\ =\ 1.10$, $p = 0.357$. According to the significant condition main effect of *RR*, the *Always Explain* condition had significantly higher *RR* than the *Never Explain* and *Pull-Prime* conditions ($p = 0.003$, $p = 0.001$, respectively; see Figure 5(a)). That is, the teams in *Always Explain* had higher coupling strength while *Never Explain* and *Pull-Prime* conditions had weaker coupling strength. Entropy findings also show that there was a significant condition main effect, $F(3,\ 60)\ =\ 3.16$, $p = 0.031$, though there was no mission main effect, $F(1,\ 60)\ =\ 0.72$, $p = 0.399$ nor an interaction effect of condition by mission, $F(3,\ 60)\ =\ 1.79$, $p = 0.158$. According to the significant condition main effect of entropy (Figure 5(b)), teams in the *Always Explain* condition had significantly higher entropy than all the other three conditions *(Never Explain*, $p = 0.019$, *Explain If Asked*, $p = 0.032$, and *Pull-Prime*, $p = 0.007$). Our prior results in Chiou et al. (2022) indicated that teams in the *Always Explain* condition also had significantly higher shared situation awareness than the other conditions, which, in light of these findings, suggests that teams in *Always Explain* had more complex in their coordination than the other three conditions, enabling them to be more adaptive to the dynamic task environment.

On the other hand, the findings for *DET* indicate that the condition main effect, $F(3,\ 60)\ =\ 1.99$, $p = 0.124$, mission main effect $F(1,\ 60)\ =\ 1.68$, $p = 0.199$, and condition by mission interaction effect, $F(3,\ 60)\ =\ 1.09$, $p = 0.361$, were not statistically significant. Similarly, *MaxL* results also show that there was no significant effects of condition, $F(3,\ 60)\ =\ 2.39$, $p = 0.078$, and mission main effects, $F(1,\ 60)\ =\ 1.63$, $p = 0.207$, nor interaction effect of condition by mission, $F(3,\ 60)\ =\ 0.74$, $p = 0.533$.

4.3. Stepwise regression

We used stepwise regression (based on Akaike information criteria, AIC) to determine the best set of predictors for shared situation awareness and team performance from the robot explanations and coordination measures. We chose stepwise regression to eliminate the *multicollinearity* issue by including an additional predictor variable and eliminating a predictor variable (i.e., forward selection and backward elimination, respectively) already in the model (Weisberg, 2005). We also chose AIC because our sample size was limited, and AIC places a moderate penalty on the number of predictor variables compared to Bayesian, which places a heavier penalty (Weisberg, 2005). This analysis was conducted in R (R

Table 3. Results for predicting shared situation awareness.

Variable	Term	*B*	*SE B*	*β*	*t*	*p*
Recurrence rate	Linear (*RR*)	−0.41	0.16	−0.23	−2.44	0.016
Explanations	Quadratic	0.01	0.01	0.33	3.54	0.001

Note. "*B*" and "*SE B*" refer to unstandardized regression coefficient and its Standard Error, respectively, while "*β*" refers to standardized regression coefficient.

Development Core Team, 2016), using the MASS packages for stepwise regression (Ripley, 2021) and lm-beta (Behrendt, 2014) for adding standardized regression coefficients.

4.4. Predicting shared situation awareness and regression diagnostics

First, we applied stepwise regression (AIC) to predict shared situation awareness via team coordination measures and explanations. The regression model accounted for 9.42% (Adjusted R-squared) of the variance ($F(2,\ 117)\ =\ 7.19$, $p = 0.001$; Table 3).

$$\widehat{Y_{TSA}} = -0.41RR + 0.01Explanations^2 + 0.87 \tag{5}$$

Based on the model shown in Equation 5, we ran regression diagnostics to investigate if the calculated model and the assumptions we made about the data and the model are consistent with the recorded data. All the assumptions were accordingly met (see Table 1); therefore, we continued summarizing the regression model in Equation 5.

According to the findings, robot explanations contributed to better shared situation awareness, and the recurrence rate adversely contributed (Table 3). We interpret this as indicating that when the robot explained the unpredictable situation, the navigator was able to code the collapses and openings on the map, which implied a better shared situation awareness. Another interesting finding is that the recurrence rate was negatively associated with shared situation awareness. These findings indicate that during the perturbations in this specific task, the robot explanations were more helpful than repetitive team coordination patterns in maintaining shared situation awareness.

4.5. Predicting team performance and regression diagnostics

We applied another stepwise regression to predict team performance by explanations and team coordination measures. The following model is obtained, $F(4,\ 115)\ =\ 3.37$, $p = 0.012$, and the model was able to account for 7.39% (adjusted R-squared) of the variance:

$$\widehat{Y_{Performance}} = -0.03MaxL + 0.04ENTR + 0.25RR + 0.74RR^2 + 1.11 \tag{6}$$

As in the previous section, we ran regression diagnostics to investigate the consistency of the calculated model and the assumptions we made about the data and the model with the recorded data. These findings are summarized in Table 2. Based on the outlier test, there were two extreme cases ($rstudent = -4.06$, $Bonferroni\ \ p = 0.011$) which we needed to exclude. However, we only excluded one of them (chosen based on the distance Q-Q plot) and re-run stepwise regression and the outlier test. Based on the new regression model, there were no extreme cases to eliminate ($rstudent = 3.32$, $Bonferroni\ \ p = 0.144$), and the rest of the regression assessment was not violated.

Table 4. Results for predicting the dyadic team performance.

Variable	Term	*B*	*SE B*	*β*	*t*	*p*
Max line	Linear (*MaxL*)	−0.04	0.01	−0.40	−2.79	0.006
Determinism	Linear (*DET*)	0.09	0.06	0.16	1.44	0.153
Recurrence rate	Linear (*RR*)	0.20	0.08	0.27	2.39	0.018
	Quadratic (*RR*^{2})	0.59	0.31	0.21	1.90	0.060

Note. "*B*" and "*SE B*" refer to unstandardized regression coefficient and its Standard Error, respectively, while "*β*" refers to standardized regression coefficient.

After the diagnostics, the final model was significant, $F(4,\ 114)\ =\ 2.64$, $p = 0.038$, and the model was able to account for 5.25% (adjusted R-squared) of the variance (see Equation 7 and Table 4):

$$\widehat{Y_{Performance}} = -0.04MaxL + 0.09DET + 0.20RR + 0.59RR^2 + 1.11 \tag{7}$$

According to the significant findings (see Table 4), stable coordination (*MaxL*) of the human-robot dyad was negatively associated with performance, but revisiting the same communication pattern (*RR*) was positively related. We interpret this as being related to how the identification of the victims might have been a more routine task in comparison to tasks that are more shared situation awareness-intensive (e.g., identifying collapses and openings); thus, this task requires more similar types of communication patterns rather than explanations (not like shared situation awareness). However, as this is still a dynamic task, and therefore, the stability (*MaxL*) in team coordination did not help for this task. This may indicate that for dyads in dynamic task environments, especially USAR or command-and-control, time pressure requires more effective interactions rather than more stable (*MaxL*) ones.

5. Discussion and conclusion

In this study, navigators made plans prior to executing their USAR missions. However, much like many real-world scenarios in which HMTs are expected to be deployed, the task environment in this study included dynamic events that required deviations from the original plan. These deviations were generally observable to the navigator but executed by the WoZ robot autonomously, and the reasons for deviations may not have been immediately apparent from the navigator's perspective. In the *Always Explain condition*, deviations from the plan were always explained by the robot proactively. This was expected to improve the navigator's understanding of the robot's behavior and improve coordination among the team members.

In dynamic task contexts, establishing shared situation awareness is a continuous process (Endsley & Jones, 2001) that involves exchanging information between teammates and coordinating team activities in response to changing external constraints. Shared situation awareness is not merely the level of shared awareness between teammates. An HMT can be treated as a single cognitive system that executes actions to perceive the environment and simultaneously act upon those perceptions towards a goal in a perception-action loop (Gorman et al., 2006).

Understanding interpersonal coordination dynamics in HMT is important for designing robotic systems that function effectively with people as teammates. This study investigated how interpersonal coordination dynamics are associated with communication strategies in a robot-assisted USAR simulation using discrete RQA. The associations between interpersonal coordination dynamics and robot explanations with shared situation awareness and team performance were also examined.

We addressed two research questions. The first question is how communication recurrence affects and reflects interpersonal coordination dynamics between the USAR robot and human navigator when using different communication strategies. The proactive provision of explanations by the robot in the *Always Explain* condition may have allowed the human navigator to adjust their behaviors in response to the robot's deviations more effectively, resulting in overall more synchronized interactions. It was also found that teams in the *Always Explain* condition had higher *ENTR* in their coordination dynamics than those in the other conditions. This indicates that the coordination dynamics were also more complex when the robot provided proactive explanations of deviations in addition to recurring proportionally more often. The greater complexity might be attributed to how the *Always Explain* condition allowed for explanations that were either proactively given by the robot or answers to questions from the navigator (i.e., more interaction types were used). This may have, in turn, led to more similar response flow patterns over time as participants learned to respond according to whether the explanation was solicited or not. The other measures of interpersonal coordination dynamics (*DET* and *MaxL*) did not differ between the conditions. Overall, these findings suggest that robot explanation appears to influence the temporal dynamics of coordination to some degree, which complements previous research that focuses on transparency, situation awareness, and trust calibration (Mercado et al., 2016).

The second research question was how the robot explanations and coordination dynamic characteristics of human-robot teams were associated with team effectiveness (i.e., shared situation awareness and performance). Team performance in this task was defined as the number of correctly triaged victims, which required locating them in the structure and marking them appropriately within the simulation. The results show that, although robot explanations improved shared situation awareness, team coordination dynamics measures were associated with team performance either positively (i.e., *RR*), negatively (i.e., *MaxL*), or not at all (i.e., *DET*, *ENTR*).

RR was positively associated with better team performance. In this task, information exchange was necessary due to role interdependence, and a stronger recurrence communication pattern is likely to reflect more effective information exchange. Although *RR* was associated with improved team performance, it was also negatively associated with shared situation awareness, indicating that teams with high shared situation awareness may still perform relatively poorly without sufficiently stable coordination dynamics. In this case, teams in *Always Explain* tended towards repeating the same communication flow patterns more often than teams in the other conditions, which helped their team performance, but not shared situation awareness.

MaxL was negatively related to team performance. *MaxL* measures the duration of the longest interaction sequences and, in this study, provided a measure of the stability and length of the HMT's coupling. This finding suggests that longer and more sustained recurrent interaction sequences were actually negatively associated with team performance. Higher *MaxL* indicates extended team interactions which may have interfered with the primary task, which needed to be interleaved with their communications. Overly stable HMT coordination dynamics have also been associated with rigidity in other dynamic team environments such as RPAS (Demir et al., 2018), which may have been detrimental to the performance given the dynamic nature of this USAR simulation. Together with the *RR* regression results, we believe this indicates that the dynamic nature of the USAR task environment requires teams to adaptively switch between different coordination patterns that are distinct from one another yet moderately stable (previously discovered by Demir, Likens, et al., 2019). Extending this further could provide evidence for multifractality in team coordination dynamics (Likens et al., 2014). This is a gap in the current studies, and it might be a future direction of this study.

In summary, the findings from this study suggest that when a robotic agent provides proactive explanations to its human navigator, it may lead to more recurrent and complex communication patterns. Altogether, our results are not surprising; repeating established communication patterns might aid in the performance of routine tasks, but they might not help as much during perturbations, in which a shared understanding of the immediate team task context is more important. Furthermore, extended and overly stable recurrent exchanges appear to be detrimental to performance in this task context. More broadly, the present results support recent efforts to understand the relationship between the temporal nature of interpersonal coordination and team effectiveness. It also demonstrates how RQA can be applied to unobtrusively increase our understanding of the temporal aspect of HMT coordination without relying on resource-consuming content-based approaches (e.g., content analysis).

5.1. Limitations and future work

There were several limitations to this study, including the generalizability of the findings from a sample of university students operating in a game-based USAR environment to designing systems that trained experts operating in real-world settings might use. To mitigate this challenge, the task environment for this study was intentionally designed to elicit key teaming and cognitive aspects of the task context from novice participants (Lematta et al., 2019). However, experience and expertise can play a large role in situation awareness. Future studies should seek to build upon our findings by examining the impact of expertise and task familiarity on shared situation awareness in the robot-teamed USAR context.

Another limitation is that though we believe that discrete RQA is sufficient for the purposes of this study, there are several types of RQA. Another RQA method to examine dyadic interaction is Cross Recurrence Quantification Analysis (CRQA; Dale et al., 2011). CRQA can also be an appropriate method for this communication flow data, because there is a binary time series for both the navigator and "robot." CRQA may have allowed us to interpret recurrence rate as coupling strength, which would provide an additional perspective into our data. Gorman et al. (2020) demonstrated that windowing could give finer-grained real-time analyses rather than analyzing the data across the trials.

Our implementation of the *Always Explain* condition was designed to promote integrated mental models in teams (Chiou et al., 2022). However, it is possible that individual differences in perceptions of robot explanations may have instead to degraded trust in the robot. For future studies, we recommend a more nuanced analysis into what constitutes sufficient or desirable levels of proactive explanation, and how matching such levels might relate to the measures explored in this study.

Finally, we acknowledge that team composition might have been a limitation in this study, as well. This task was based upon key elements of real-world USAR operations combined with WoZ-enabled robot behaviors to answer research questions about human-robot interactions. However, real USAR operations typically take place as part of larger teams (Murphy, 2019). Future work can extend these results to teams with more than two members to see how interpersonal coordination affects team effectiveness within larger and multi-level teams.

Acknowledgments

We acknowledge the assistance of Steven M. Shope and Paul Jorgenson, Sandia Research Corporation, who developed and provided the chat system. We also thank the many CERTT Lab students for their assistance in developing the task environment, data collection, and pre-processing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Air Force Office of Scientific Research [FA9550-18–1-0067]. Any findings or recommendations expressed in this material are those of the authors and do not necessarily reflect the views or policies of the research sponsor.

ORCID

Mustafa Demir http://orcid.org/0000-0002-5667-3701

References

Abarbanel, H. (1996). *Analysis of observed chaotic data* (1st ed.). Springer.

Allen, L. K., Perret, C., Likens, A., & McNamara, D. S. (2017). What'D you say again?: Recurrence quantification analysis as a method for analyzing the dynamics of discourse in a reading strategy tutor. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 373–382). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027445

Arthur, W. B., Beinhocker, E., & Stanger, A. (Eds.). (2020). *Complexity economics: Proceedings of the Santa Fe Institute's 2019 fall symposium*. SFI Press.

Bartlett, C. E., & Cooke, N. J. (2015). Human-robot teaming in urban search and rescue. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 250–254. https://doi.org/10.1177/1541931215591051

Behrendt, S. (2014, December 28). *Add standardized regression coefficients to lm-objects [R package lm.beta version 1.5-1]*. Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=lm.beta

Brannick, M. T., Prince, A., Prince, C., & Salas, E. (1995). The measurement of team process. *Human Factors*, *37*(3), 641–651. https://doi.org/10.1518/001872095779049372

Butler, E. A. (2011). Temporal interpersonal emotion systems: The "TIES" that form relationships. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *15*(4), 367–393. https://doi.org/10.1177/1088868311411164

Butner, J. E., Berg, C. A., Baucom, B. R., & Wiebe, D. J. (2014). Modeling coordination in multiple simultaneous latent change scores. *Multivariate Behavioral Research*, *49*(6), 554–570. https://doi.org/10.1080/00273171.2014.934321

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo, E. Salas, & Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 333–380). Jossey-Bass.

Cappella, J. N. (1987). Interpersonal communication: Definitions and fundamental questions. In C. R. Berger & S. H. Chaffee (Eds.), *Handbook of communication science* (pp. 184–238). De Gruyter Mouton.

Chanel, G., Bétrancourt, M., Pun, T., Cereghetti, D., & Molinari, G. (2013). Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling. In *Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 116–122). IEEE. https://doi.org/10.1109/ACII.2013.26

Chiou, E. K., Demir, M., Buchanan, V., Corral, C. C., Endsley, M. R, Lematta, G. J., Cooke, N. J., & McNeese, N. J. (2022). Towards human–robot teaming: Tradeoffs of explanation-based communication strategies in a virtual search and rescue task. *International Journal of Social Robotics*, *14*, 1117–1136. https://doi.org/10.1007/s12369-021-00834-1

Chiou, E. K., & Lee, J. D. (2016). Cooperation in human-agent systems to support resilience: A microworld experiment. *Human Factors*, *58*(6), 846–863. https://doi.org/10.1177/0018720816649094

Chiou, E. K., & Lee, J. D. (2021). Trusting automation: Designing for responsivity and resilience. *Human Factors*. https://doi.org/10.1177/00187208211009995

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (1st ed.). Academic Press. https://www.elsevier.com/books/statistical-power-analysis-for-the-behavioral-sciences/cohen/978-0-12-179060-8

Cohen, M. C., Demir, M., Chiou, E. K., & Cooke, N. J. (2021, September 15). The dynamics of trust and verbal anthropomorphism in human-autonomy teaming. In *2nd IEEE International Conference on Human-Machine Systems (ICHMS)* (pp. 1–6). https://doi.org/10.1109/ICHMS53169.2021.9582655

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Journal of Cognitive Engineering and Decision Making*, *3*(1), 27–46. https://doi.org/10.1518/155534309X433302

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

Cummings, M. L. (2006). Automation and accountability in decision support system interface design. *Journal of Technology Studies*, *32*(1), 23–31. https://doi.org/10.21061/jots.v32i1.a.4

Dale, R., Warlaumont, A. S., & Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos*, *21*(04), 1153–1161. https://doi.org/10.1142/S0218127411028970

de Visser, E. J., Peeters, M. M. M., Jung, M. F., Kohn, S., Shaw, T. H., Pak, R., & Neerincx, M. A. (2020). Towards a theory of longitudinal trust calibration in human–robot teams. *International Journal of Social Robotics*, *12*(2), 459–478. https://doi.org/10.1007/s12369-019-00596-x

Demir, M., Amazeen, P. G., & Cooke, N. J. (2020). Examining human-autonomy team interaction and explicable behavior in a dynamic LEGO construction task. *Procedia Computer Science*, *168*, 195–201. https://doi.org/10.1016/j.procs.2020.02.270

Demir, M., Cooke, N. J., & Amazeen, P. G. (2018). A conceptual model of team dynamical behaviors and performance in human-autonomy teaming. *Cognitive Systems Research*, *52*, 497–507. https://doi.org/10.1016/j.cogsys.2018.07.029

Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2019). Team coordination and effectiveness in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems*, *49*(2), 150–159. https://doi.org/10.1109/THMS.2018.2877482

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3–12. https://doi.org/10.1016/j.cogsys.2016.11.003

Demir, M., McNeese, N. J., & Cooke, N. J. (2018a). The impact of perceived autonomous agents on dynamic team behaviors. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *2*(4), 258–267. https://doi.org/10.1109/TETCI.2018.2829985

Demir, M., McNeese, N. J., & Cooke, N. J. (2018b). Dyadic team interaction and shared cognition to inform human-robot teaming. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 124–124. https://doi.org/10.1177/1541931218621028

Demir, M., McNeese, N. J., & Cooke, N. J. (2019). The evolution of human-autonomy teams in remotely piloted aircraft systems operations. *Frontiers in Communication*, *4*, 1–12. https://doi.org/10.3389/fcomm.2019.00050

Demir, M., McNeese, N. J., & Cooke, N. J. (2020). Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies*, *140*, 102436. https://doi.org/10.1016/j.ijhcs.2020.102436

Demir, M., McNeese, N. J., Gorman, J. C., Cooke, N. J., Myers, C. W., & Grimm, D. A. (2021). Exploration of team trust and interaction dynamics in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems*, *51*(6), 696–705. https://doi.org/10.1109/THMS.2021.3115058

Eckmann, J.-P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters (EPL)*, *4*(9), 973–977. https://doi.org/10.1209/0295-5075/4/9/004

Endsley, M., & Jones, W. (2001). A model of inter- and intra-team situation awareness: Implications for design, training and measurement. In M. McNeese, E. Salas, & M. Endsley (Eds.), *New trends in cooperative activities: Understanding system dynamics in complex environments* (pp. 46–67). Human Factors and Ergonomics Society.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiore, S. M., & Wiltshire, T. J. (2016). Technology as teammate: Examining the role of external cognition in support of team cognitive processes. *Frontiers in Psychology*, *7*, 1–17. https://doi.org/10.3389/fpsyg.2016.01531

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere, R., Maechler, M., Monette, G., Murdoch, D., … Nilsson, H. (2022). *car: Companion to Applied Regression* (3.0-13) [Computer software]. https://CRAN.R-project.org/package=car

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273–286. https://doi.org/10.1016/j.chb.2004.02.014

Gorman, J., Amazeen, P., & Cooke, N. (2010). Team coordination dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences*, *14*(3), 265–289.

Gorman, J. C., Cooke, N. J., Amazeen, P. G., & Fouse, S. (2012). Measuring patterns in team interaction sequences using a discrete recurrence approach. *Human Factors*, *54*(4), 503–517. https://doi.org/10.1177/0018720811426140

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, *49*(12–13), 1312–1325. https://doi.org/10.1080/00140130600612788

Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring real-time team cognition during team training. *Human Factors*, *62*(5), 825–860. https://doi.org/10.1177/0018720819852791

Gorman, J. C., Hessler, E. E., Amazeen, P. G., Cooke, N. J., & Shope, S. M. (2012). Dynamical analysis in real time: Detecting perturbations to team communication. *Ergonomics*, *55*(8), 825–839. https://doi.org/10.1080/00140139.2012.679317

Gorman, J. C., Martin, M. J., Dunbar, T. A., Stevens, R. H., Galloway, T. L., Amazeen, P. G., & Likens, A. D. (2016). Cross-level effects between neurophysiology and communication during team training. *Human Factors*, *58*(1), 181–199. https://doi.org/10.1177/0018720815602575

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

Guastello, S. J. (2010). Nonlinear dynamics of team performance and adaptability in emergency response. *Human Factors*, *52*(2), 162–172. https://doi.org/10.1177/0018720809359003

Guastello, S. J., & Guastello, D. D. (1998). Origins of coordination and team effectiveness: A perspective from game theory and nonlinear dynamics. *Journal of Applied Psychology*, *83*(3), 423–437. https://doi.org/10.1037/0021-9010.83.3.423

Guastello, S. J., & Peressini, A. F. (2017). Development of a synchronization coefficient for biosocial interactions in groups and teams. *Small Group Research*, *48*(1), 3–33. https://doi.org/10.1177/1046496416675225

Gudykunst, W. B. (2000). *Communication yearbook*. SAGE.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Mashkati (Eds.), *Human mental workload* (pp. 139–183). North Holland Press.

Hornyak, T. (2020, March 18). What America can learn from China's use of robots and telemedicine to combat the coronavirus. *CNBC*. https://www.cnbc.com/2020/03/18/how-china-is-using-robots-and-telemedicine-to-combat-the-coronavirus.html

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems.

International Journal of Cognitive Ergonomics, *4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Johnson, C. J., Demir, M., McNeese, N. J., Gorman, J. C., Wolff, A. T., & Cooke, N. J. (2021). The impact of training on human–autonomy team communications and trust calibration. *Human Factors*. https://doi.org/10.1177/00187208211047323

Kantz, H., & Schreiber, T. (1997). *Nonlinear time series analysis* (1 ed.). Cambridge University Press.

Kelley, J. F. (1983, December). An empirical methodology for writing user-friendly natural language computer applications. In *Conference on Human Factors in Computing Systems - Proceedings* (pp. 193–196). Association for Computing Machinery. https://doi.org/10.1145/800045.801609

Kopp, T., Baumgartner, M., & Kinkel, S. (2022). How linguistic framing affects factory workers' initial trust in collaborative robots: The interplay between anthropomorphism and technological replacement. *International Journal of Human-Computer Studies*, *158*, 102730. https://doi.org/10.1016/j.ijhcs.2021.102730

Lematta, G. J., Coleman, P. B., Bhatti, S. A., Chiou, E. K., McNeese, N. J., Demir, M., & Cooke, N. J. (2019). Developing human-robot team interdependence in a synthetic task environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 1503–1507. https://doi.org/10.1177/1071181319631433

Likens, A. D., Amazeen, P. G., Stevens, R., Galloway, T., & Gorman, J. C. (2014). Neural signatures of team coordination are revealed by multifractal analysis. *Social Neuroscience*, *9*(3), 219–234. https://doi.org/10.1080/17470919.2014.882861

Lin, L., & Xu, C. (2020). Arcsine-based transformations for meta-analysis of proportions: Pros, cons, and alternatives. *Health Science Reports*, *3*(3), e178. https://doi.org/10.1002/hsr2.178

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426. https://doi.org/10.1111/j.1551-6709.2012.01269.x

Maes, P. (1993). Modeling adaptive autonomous agents. *Artificial Life*, *1*(1_2), 135–162. https://doi.org/10.1162/artl.1993.1.1_2.135

Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, *26*(1), 87–119. https://doi.org/10.1145/174666.174668

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *The Academy of Management Review*, *26*(3), 356–376. https://doi.org/10.5465/amr.2001.4845785

mars.nasa.gov (n.d). *Goals*. Retrieved February 25, 2021, from https://mars.nasa.gov/mars2020/mission/science/goals/

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5–6), 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

Marwan, N., & Webber, C. L. (2015). Mathematical and computational foundations of recurrence quantifications. In C. L. Webber, & N. Marwan (Eds.), *Recurrence quantification analysis* (pp. 3–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-07155-8_1

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for multi-UxV management. *Human Factors*, *58*(3), 401–415. https://doi.org/10.1177/0018720815621206

Miller, C. A. (2021). Trust, transparency, explanation, and planning: Why we need a lifecycle perspective on human-automation interaction. In *Trust in human-robot interaction* (pp. 233–257). Elsevier. https://doi.org/10.1016/B978-0-12-819472-0.00011-3

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Murphy, R. R. (2019). *Introduction to AI robotics*. MIT Press.

Nalepka, P., Gregory-Dunsmore, J. p., Simpson, J., Patil, G., & Richardson, M. (2021, May 13). Interaction flexibility in artificial agents teaming with humans. *Proceedings of the Annual Meeting of the Cognitive Science Society, 43*. Retrieved from https://escholarship.org/uc/item/9ks6n70q

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2022). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *64*(5), 904–938. https://doi.org/10.1177/0018720820960865

Open AI, Berner, C., Brockman, G., Chan, B., Cheung, V., Dȩbiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., Pinto, H. P. D O., Raiman, J., Salimans, T., Zhang, S. … (2019). Dota 2 with large scale deep reinforcement learning. ArXiv191206680 [CsStat]. http://arxiv.org/abs/1912.06680

Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical Review Letters*, *45*(9), 712–716. https://doi.org/10.1103/PhysRevLett.45.712

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *21*(2), 99–141. https://doi.org/10.1177/1088868316628405

Pearce, W. B. (1976). The coordinated management of meaning: A rules-based theory of interpersonal communication. In *Explorations in interpersonal communication* (pp.278–278). Sage.

Pellerin, C. (2015). *Work: Human-machine teaming represents defense technology future*. U.S. Department of Defense. https://www.defense.gov/Explore/News/Article/Article/628154/work-human-machine-teaming-represents-defense-technology-future/

Pena, E. (2019). *Package 'gvlma': Global validation of linear models assumptions* (1.0.0.3) [R Software]. CRAN. https://cran.r-project.org/web/packages/gvlma/gvlma.pdf

Perone, S., & Simmering, V. R. (2017). Applications of dynamic systems theory to cognition and development: New frontiers. In J. B. Benson (Ed.), *Advances in child development and behavior* (Vol. 52, Chapter 2, pp 43–80). Academic Press.

R Development Core Team. (2016, October 31). *R: A programming environment for data analysis and graphics (Version 3.3.2)*. https://www.r-project.org/

Ramos-Villagrasa, P. J., Marques-Quinteiro, P., Navarro, J., & Rico, R. (2018). Teams as complex adaptive systems: Reviewing 17 years of research. *Small Group Research*, *49*(2), 135–176. https://doi.org/10.1177/1046496417713849

Ripley, B. (2021, May 3). *Support functions and datasets for Venables and Ripley's MASS [R package MASS version* 73-54]. Comprehensive R Archive Network (CRAN). https://CRAN.R-project.org/package=MASS

Romero, M. E. (2020, April 8). Tommy the robot nurse helps Italian doctors care for COVID-19 patients. *The World from PRX*. https://www.pri.org/stories/2020-04-08/tommy-robot-nurse-helps-italian-doctors-care-covid-19-patients

Russell, S. M., Funke, G. J., Knott, B. A., & Strang, A. J. (2012). Recurrence quantification analysis used to assess team communication in simulated air battle management. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 468–472. https://doi.org/10.1177/1071181312561046

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, *50*(3), 540–547. https://doi.org/10.1518/001872008X288457

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 123–136. https://doi.org/10.1518/001872095779049525

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In *Coordination: Neural, Behavioral and Social Dynamics* (pp. 281–308). Springer. https://doi.org/10.1007/978-3-540-74479-5_14

Schultz, D., Spiegel, S., Marwan, N., & Albayrak, S. (2015). Approximation of diagonal line based measures in recurrence quantification analysis. *Physics Letters A*, *379*(14–15), 997–1011. https://doi.org/10.1016/j.physleta.2015.01.033

Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human Factors*, *52*(2), 234–245. https://doi.org/10.1177/0018720809350882

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples)†. *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Spencer, J. P., & Perone, S. (2008). Defending qualitative change: The view from dynamical systems theory. *Child Development*, *79*(6), 1639–1647. https://doi.org/10.1111/j.1467-8624.2008.01214.x

Stergiou, N. (2016). *Nonlinear analysis for human movement variability*. CRC Press.

Stowers, K., Brady, L. L., MacLellan, C., Wohleber, R., & Salas, E. (2021). Improving teamwork competencies in human-machine teams: Perspectives from team science. *Frontiers in Psychology*, *12*, 1–6. https://doi.org/10.3389/fpsyg.2021.590290

Strang, A. J., Funke, G. J., Russell, S. M., Dukes, A. W., & Middendorf, M. S. (2014). Physio-behavioral coupling in a cooperative team task: Contributors and relations. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(1), 145–158. https://doi.org/10.1037/a0033125

Thelen, E., & Smith, L. B. (2007). Dynamic systems theories. In *Handbook of child psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0106

Weisberg, S. (2005). Variable selection. In *Applied linear regression* (pp. 211–232). John Wiley & Sons, Inc. https://doi.org/10.1002/0471704091.ch10/summary

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-solving phase transitions during team collaboration. *Cognitive Science*, *42*(1), 129–167. https://doi.org/10.1111/cogs.12482

Wiltshire, T. J., Philipsen, J. S., Trasmundi, S. B., Jensen, T. W., & Steffensen, S. V. (2020). Interpersonal coordination dynamics in psychotherapy: A systematic review. *Cognitive Therapy and Research*, *44*(4), 752–773. https://doi.org/10.1007/s10608-020-10106-3

About the authors

Mustafa Demir is an assistant research professor in Global Security Initiative at Arizona State University. Dr. Demir received his Ph.D. in Simulation, Modelling, and Applied Cognitive Science with a focus on team coordination dynamics and effectiveness in human-machine teaming from Arizona State University in Spring 2017.

Myke Cohen is a Ph.D. student in Human Systems Engineering and Ira A. Fulton Schools of Engineering Dean's Fellow at Arizona State University. His research center on socio-temporal aspects of teaming in complex work environments. He holds a B.S. in Industrial Engineering from the University of the Philippines, Diliman.

Craig J. Johnson is a Human Systems Engineering Ph.D. candidate and Ira A. Fulton Schools of Engineering Dean's Fellow at Arizona State University. He holds a Bachelor of Science in psychology from Clemson University and a Master of Science in Human Systems Engineering from Arizona State University.

Erin K. Chiou is an assistant professor of human systems engineering at Arizona State University and directs the Automation Design Advancing People and Technology (ADAPT) Laboratory. She received her Ph.D. in industrial engineering from the University of Wisconsin-Madison, and B.S. in psychology and philosophy from the University of Illinois, Urbana-Champaign.

Nancy J. Cooke is a professor of Human Systems Engineering at Arizona State University and directs ASU's Center for Human, Artificial Intelligence, and Robot Teaming. Dr. Cooke studies individual and team cognition and its application to human, AI, and robot teaming and conducts empirical assessments of teams and teamwork.