Exploration of Teammate Trust and Interaction Dynamics in Human-Autonomy Teaming

Mustafa Demir, *Member, IEEE*, Nathan J. McNeese, *Member, IEEE*, Jaime C. Gorman, Nancy J. Cooke, Christopher W. Myers, and David A. Grimm

***Abstract*—This article considers human-autonomy teams (HATs) in which two human team members interact and collaborate with an autonomous teammate to achieve a common task while dealing with unexpected technological failures that were imposed either in automation or autonomy. A Wizard of Oz methodology is used to simulate the autonomous teammate. One of the critical aspects of HAT performance is the trust that develops over time as team members interact with each other in a dynamic task environment. For this reason, it is important to examine the dynamic nature of teammate trust through real-time measures of team interactions. This article examines team interaction and trust to understand better how they change under automation and autonomy failures. Thus, we address two research questions: 1) How does trust in HATs evolve over time?; and 2) How is the relationship between team interaction and trust impacted by the failures? We hypothesize that trust in HATs will decrease as autonomy failures increase. We also hypothesize that team interaction would be related to the development of trust and recovery from the failures. The results implicate three general trends: 1) team interaction dynamics are linked to the development of trust in HATs; 2) trust in the autonomous teammate is only associated with recovery from autonomy failures; 3) team interaction dynamics are related to both automation and autonomy failure recovery.**

***Index Terms*—Artificial intelligence, dynamical systems, team coordination, trust, unmanned air vehicle systems.**

I. Introduction

THANKS to the advancement of computational algorithms, autonomous agents have become intelligent enough to be considered a teammate as opposed to a tool [1]. This advancement is also reflected in the scientific terminology by shifting from "automation" to "autonomy." Automation is defined as: "1) the mechanization and integration of the sensing of environmental variables (by artificial sensors); 2) data processing and decision making (by computers); and 3) mechanical action (by motors or devices that apply forces on the environment)" [2, p. 9]. In contrast, autonomy is a machine that can carry out tasks independently and in conjunction with human interactions [3]. In this article, we focus on autonomy as a teammate that works with others to control automation, a remotely piloted aircraft. Autonomy is an intelligent machine (e.g., artificial intelligence - AI, robot, synthetic agent), considered a teammate [3]. Autonomous agents have gotten better at incorporating feedback and information to make plans and take action towards their goals [4]. This article also considers human-autonomy teams (HATs) as teams that include at least one heterogeneous and interdependent autonomous machine team member. In HATs, human and autonomous teammates promptly interact with one another in response to information flow from one team member to another, adapt to the dynamic task, and achieve common goals [5].

Automation research focuses on the effects of automation on user performance and subjective perception. For autonomy, however, the focus is shifted to interactions (i.e., communication and coordination) between human and autonomous team members, and in turn, emotional reactions to the autonomy, such as trust. Consequently, social and psychological factors are considered to a greater extent in autonomy than automation [6]. In HAT, one of the critical psychological aspects is trust that is defined as "the extent to which a person is confident in and willing to act on the basis of the words, actions, and decisions of another" [7]. Trust in HAT is a personal belief that an autonomous agent with which one is coordinating can achieve a common goal in a task environment [8]. It is critical to understand that trusting an autonomous agent is not only trusting a machine but also trusting it as a team member [9], which requires several beliefs about others [10]. Specifically, teammate trust, also called mutual trust, is defined as "the shared belief that team members will perform their roles and protect the interests of their teammates," which also leads to a "willingness to admit mistakes and accept feedback" [11]. Other team-related trust studies align trust with "vulnerability" [13], [14], and thus, a willingness to take risks [13]. Being vulnerable means that something important can be lost, and making oneself vulnerable is taking a risk. There is an inherent risk in dynamic, team-based

Manuscript received October 26, 2019; revised April 30, 2020, January 18, 2021, and July 3, 2021; accepted September 19, 2021. Date of publication October 7, 2021; date of current version November 12, 2021. This work was supported by the Office of Naval Research under Grant N000141712382. This article was recommended by Associate Editor K. M. Feigh. *(Corresponding author: Mustafa Demir.)*

Mustafa Demir and Nancy J. Cooke are with the Human Systems Engineering, Ira A Fulton Schools of Engineering, Arizona State University, Tempe, AZ 85281 USA (e-mail: mdemir@asu.edu; ncooke@asu.edu).

Nathan J. McNeese is with the Human-Centered Computing, Clemson University, Clemson, SC 29634 USA (e-mail: mcneese@clemson.edu).

Jaime C. Gorman is with the Industrial Psychology, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: jamie.gorman@psych.gatech.edu).

Christopher W. Myers is with the Cognitive Science, Models, Agents Branch, Wright-Patterson AFB, Dayton, OH 45433 USA (e-mail: christopher.myers.29@us.af.mil).

David A. Grimm is with the Engineering Psychology, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: david.grimm@gatech.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2021.3115058.

Digital Object Identifier 10.1109/THMS.2021.3115058

2168-2291 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

tasks attributable to team members' role interdependency to accomplish a common goal. For this article, we include questions regarding human team members' willingness to take a risk [13] and commonly used trust and reliability questions in robotics questionnaires [14].

De Visser *et al.* [15] developed an interactive model related to trust calibration (reducing overtrust and increasing undertrust) and its impact on human-robot team effectiveness. Based on their interactive model, it is crucial to understand the relationship between trust calibration and team interaction behaviors and their relation to team effectiveness. Other studies regarding trust also emphasize that trust is an interactive process with multiple influences. In a multiagent team, getting an evidence-based weighted average of trust in each agent from multiple directly involved humans may help identify under what circumstances and contexts autonomous agents tend to fail or perform below the desired threshold (e.g., human capabilities) [16]. However, if trust is dynamic and emerges through interaction, these ratings should be predicted by interaction-based real-time physiological and psychological measures. Therefore, in addition to subjective measures, real-time objective measures of interaction should also be considered when assessing trust in HATs.

In order to predict the dynamics of teammate trust, some studies have considered real-time measures, including human teammates' physiological processes, such as heart rate [9] and interaction between team members in all-human team and HAT contexts [19], [20]. Many of these studies use nonlinear dynamical systems analytical methods, specifically, recurrence quantification analysis (RQA; [21], [22]). A multivariate extension of RQA investigates interaction patterns between the system components (in this case, either the heart rate of team members or their communication flow) and their change over time. Mitkidis *et al.* [17] studied how trust modulates the affective links between team members by applying multivariate RQA to assess the degree of synchrony in dyadic all-human teams. In general, team synchrony occurs when two or more systems (or two individuals in a team) are in behavioral sync, resulting in their recurrent behaviors being dependent on each other. Mitkidis *et al.* found that trust was associated with heart rate synchrony, such that interpersonal physiological synchrony may be an indicator of interpersonal trust. Tolson *et al.* [18] also applied the multivariate extension of RQA to examine how interpersonal physiological arousal predicts willingness to trust an autonomous agent in unexpected conditions for two- or three-member teams. Their findings indicate that interpersonal team-level arousal is a significant predictor of trusting an autonomous agent during unexpected events. Another study by Grimm *et al.* [19] applied multivariate RQA on team interaction to understand if communication between team members in HATs was synchronized in a remote pilot aircraft system. Their study increased the workload by imposing technological failures. Results indicated that human team members' trust and reliance on their autonomous team member remained relatively constant over time within low-performing teams. However, in high-performing teams, human team members' trust and reliance on the autonomous team member diminished.

Failures of automation and autonomy can adversely affect the operator's trust [8] and system performance [24], [25]. However, when the automation and autonomy design is more adaptive (i.e., adaptive automation), trust increases, and task workload decreases, resulting in an increment of team performance [25]. One aim of this article is to elucidate further the relationship between team interaction dynamics and trust to better articulate how they change in the context of automation and autonomy failures. Overall, this aim underlines trust as an evolving process that occurs when human and autonomous team members interact with each other and adapt to their respective capabilities within a dynamic task environment.

Consequently, it is important to establish relationships between subjective trust measures and other real-time objective measures, including communication, coordination, and physiological measures. We report results from a study investigating trust in the context of a simulated remotely-piloted aircraft system (RPAS) task environment. Specifically, we examine interaction dynamics and perceived trust to understand better how they change over time in the context of RPAS autonomy and automation failures. Thus, we address the following two research questions: 1) How does trust in HATs evolve over time?; and 2) How is the relationship between team interaction and trust impacted by failures? We hypothesized that trust in HATs would decrease as autonomy failures increased. We also hypothesized that team interaction would be related to the development of trust and recovery from failures. We hypothesized that trust would decrease as autonomy failures increase across experimental sessions. We also hypothesized that team interaction would be related to trust development and automation and autonomy failure recovery.

II. CURRENT STUDY

A. Task Environment

The task environment called the Cognitive Engineering Research on Team Tasks RPAS Synthetic Task Environment (CERTT-RPAS-STE) comprises three task roles: communicating via text chat. This RPAS-STE was utilized to simulate aspects of the MQ-1 predator remotely piloted aircraft (RPA), an aircraft used by the United States Air Force for aerial reconnaissance military operations related to teamwork [25]. The objective of the RPAS-STE was to take photographs of color-coded strategic target waypoints. Three individual, yet interdependent, team members cooperate to accomplish this goal: 1) a *navigator*, who has access to data regarding the location of and any restrictions on critical waypoints; using this information, the navigator plans a sequence of waypoints called the mission route and notifies the pilot regarding the waypoints, including waypoint name, altitude restrictions, airspeed restrictions, and effective target radius; 2) a *pilot*, who controls and monitors the altitude of the RPA, airspeed, effective radius of the current waypoint, fuel, gears, and flaps, and also interacts with the photographer to negotiate regarding altitude and airspeed to obtain a good photograph of the target waypoint; and 3) a *photographer*, who monitors and adjusts camera settings to take the photos of targets and then sends feedback to the other teammates regarding photo quality.

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

To achieve this, the photographer needs to monitor and control the systems and settings relevant to the RPA sensor equipment (i.e., infrared and electro-optical photography equipment), and coordinate with the pilot about the right altitude and airspeed to take a good photo. In this task, the coordination of communication between the team members is as follows: the navigator sends the waypoint *information* to the pilot. Then, the pilot *negotiates* the altitude of the waypoint with the photographer who needs the correct altitude to take a good photo of the waypoint. Finally, the photographer takes a good photo and sends *feedback* to both team members [26].

B. Design

This article followed a "*Wizard of Oz*" (WoZ [27]) paradigm wherein the navigator and photographer were seated together in one room and were told that the pilot was an autonomous agent. In actuality, the pilot was a well-trained experimenter who was working from a separate room. The "autonomous" pilot, who used restricted vocabulary and a predetermined script to simulate that of an autonomous agent, interacted with the other teammates in a timely manner, similar to a software-based autonomous pilot used in a previous experiment [28]. Due to the autonomous pilot's limited language capabilities, "cheat sheets" were provided to the human teammates to be used during training and task performance to assist in effective communication with the pilot. The main manipulation consisted of three failures as follows:

1) *automation*- role-specific display failures that occurred while processing specific targets;
2) *autonomy*- autonomous agent behaved abnormally while processing specific targets (i.e., it provided misinformation to other team members or demonstrated incorrect action);
3) c*yber attack*- the hijacking of the pilot, which led to the autonomous pilot providing detrimental information to the team [29].

In the previous study [29], we found that team interaction between all three team members was necessary for the automation failure recovery, because when a display fails, team members (either human or autonomous teammate) exchanged the information regarding the target waypoint. In turn, they were able to take a good photo. However, the autonomy failure and malicious cyber-attack recoveries required continuous and persistent information given to the autonomous pilot from the human team members to correct its abnormal behavior. The current experiment was comprised of ten 40-min missions. Each failure was imposed on a preselected target waypoint (see Table I). Automation and autonomy failures were each imposed once during each mission, beginning with Mission 2. In contrast, the malicious cyber-attack was imposed only once in Mission 10, and the teams had to find a solution in a limited amount of time to recover from these failures and take a good photo. The time limit for each failure was related to the failure's difficulty and was determined to be around 400 s, based on pilot testing. Because the cyber attack only occurred once, we have focused on automation and autonomy failures.

TABLE I
FAILURE TYPES PER MISSION

		Automation	**Autonomy**	**Cyber Attack**
Session I	Training	No Failure	No Failure	No Failure
	Mission 1	No Failure	No Failure	No Failure
	Mission 2	2nd target	4th target	No Failure
	Mission 3	4th target	2nd target	No Failure
	Mission 4	1st target	3rd target	No Failure
Session II	Mission 5	2nd target	4th target	No Failure
	Mission 6	4th target	2nd target	No Failure
	Mission 7	1st target	3rd target	No Failure
	Mission 8	3rd target	1st target	No Failure
	Mission 9	3rd target	5th target	No Failure
	Mission 10	2nd target	4th target	Last 10 minutes

Note: Between the two sessions, there was a one- or two-week interval. A 15-min break was given after each task; a half-hour lunch break was given.

III. Methodology

A. Participants

In total, 22 teams (44 participants) were recruited for participation from Arizona State University and the surrounding community; all teams completed the experiment. Two participants per team were randomly assigned to the photographer and the navigator roles, and the pilot position was filled by a well-trained experimenter who mimicked an autonomous pilot in terms of communication and coordination [28], the *WoZ autonomous pilot*. Participation required normal or corrected-to-normal vision and fluency in English. Participants ranged in age from 18 to 36 ($M_{\text{age}} = 23$, $SD_{\text{age}} = 3.90$) with 21 males and 23 females and were undergraduate or graduate students. Each team participated in two seven-hour sessions, and each participant was compensated for participation by payment of \$10 per hour. This article was carried out according to The Cognitive Engineering Research Institute (CERI) Institutional Review Board under the CERI. The protocol was approved by the Cognitive Engineering Research Institute Institutional Review Board. All subjects gave written informed consent by the Declaration of Helsinki.

B. Materials and Apparatus

The CERTT-RPAS-STE hardware consisted of four consoles for up to four team members and another four consoles for two experimenters to oversee the simulation, apply failures, and observe the team. Two of the consoles were actively used by two experimenters, and the other two were used for data storage. One of the actively used consoles was called the "texting experimenter" console. This console had two computers through which the experimenter could make ratings (for target achievement, behavior, and situation awareness) and send text messages via chat to the three teammates (pilot, navigator, and photographer). This console was also used to turn all the computers and software in the other consoles on and off via a master control application. Another actively used console was the "nontexting experimenter" console, which enabled the experimenter to make ratings (for coordination, target achievement, and behavior) and monitor the participants using a camera. In addition, text

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

chat capability was provided for communications between team members.

For this article, there were other materials and equipment used to collect the measures and train participants. For hands-on training and the task, the photographer and the navigator had a communication "cheat sheet," which showed examples of how to communicate with the autonomous pilot, which was necessary due to its limited communication capability. Another "cheat sheet" and supplemental materials for each role were displayed at the corresponding workstations. The pilot role in the experimenter condition had a coordination script in order to push and pull information to and from other team members in a timely manner. Supplemental materials included role-specific rule summaries, screenshots of each station's displays, a waypoint list for the navigator, and a camera setting list and photo folder containing comparisons of good and bad photos for the photographer. Experimenters followed paper checklists for the experiment set-up (starting the experimenter and participant consoles, briefing, training, and mission task), and data for each session was archived on an external hard drive.

C. Experimental Procedure and Measures

The experiment was divided into two seven-hour sessions with a one- or two-week interval between the sessions. When participants arrived, they read and signed an informed consent form and were randomly assigned to one of the two team member roles, navigator or photographer; the pilot was an expert confederate who used a role-specific coordination script. The pilot was isolated in one room and the navigator and photographer in another. The navigator and photographer were seated in locations separated by partitions.

Participants then received a briefing followed by 30-min of role-specific skills training using interactive PowerPoint slides. After the interactive training session, 30-min of hands-on practice training began. During the practice session, the experimenters used a checklist to ensure that the navigator and photographer were comfortable performing their roles. Once experimenters were sure that the participants understood their individual and team tasks, they started the first mission. Immediately after the practice training and each mission, participants were shown their performance scores. Participants could see their team performance score and each individual's task-related individual performance score. The performance scores were displayed on each participant's computer and shown in comparison to the mean scores achieved by all other teams (or roles) who had participated in the experiment up to that point. After Missions 1, 4, 5, and 10, the NASA-TLX subjective workload assessment [30] was administered to the navigator and photographer to measure six workload components: mental, physical, temporal demand, performance, effort, and frustration. Throughout the missions, team communication was observed by texting and nontexting experimenters who checked appropriate boxes on situation awareness and coordination loggers in real-time. After visiting each target waypoint, both experimenters independently rated the team's process behaviors on a scale from one to five, five being the best. A short set of anthropomorphism- and trust-related questions [13] was administered to the navigator and photographer after Missions 4 and 10. A demographic questionnaire was administered after Mission 10. At the end of Session II, a debriefing was provided.

In this experiment, three-team performance measures were obtained as follows:

1) a mission level performance score (a composite score calculated based on the overall RPAS team);
2) target processing efficiency (based on the timely and accurate processing of the target);
3) whether or not the teams successfully recovered after the failures.

In addition, this article aimed to highlight and discuss trust and team interaction mechanisms to help recover from complex and dynamic failures. Hence, we focus on the following measures.

Team communication flow was used to quantify team interaction dynamics measures, and is hypothesized to be related to teammate trust. Team communication flow is a multivariate binary measure recorded once each minute for each team member to indicate if at least one message was sent (team communication flow = 1) or not (team communication flow = 0) by each team member.

Trust was measured subjectively by giving a questionnaire to the navigator and photographer roles. We considered the study done by Mayer and Gavin (2005) because of our main point of interest, i.e., teams. The wording of Mayer and Gavin's questionnaire [13], which contains "willingness to take a risk" questions relating to vulnerability and also included additional trust questions, was modified for use in this HAT context [31]. The questionnaire had 25 questions with a Likert scale ranging from "1" = Strongly Agree to "5" = Strongly disagree. To assess how teammate trust changed over time, the questionnaire was administered after Sessions I and II.

Automation and autonomy failure recovery scores were separately calculated for Sessions I and II by taking the proportion of times that either the automation or autonomy failures were successfully resolved in each session.

IV. Data Analytics and Results

The following analyses were applied to understand the impact of different characteristics of team interaction dynamics on human team members' trust in the autonomous pilot. First, exploratory factor analysis (EFA) was applied to the 25-item trust questionnaire to uncover the underlying factor structure, i.e., factor scores. Then, Joint Recurrence Quantification Analysis (JRQA) was applied to the multivariate team communication flow to calculate the interaction dynamics measures: recurrence rate (RR), percent determinism (DET), and maximum line length (MaxL) in each mission [22]. We averaged the interaction dynamics measures (DET, RR, and MaxL) across each session to equalize the data length of trust and team interaction measures. Z-scores (i.e., standard scores) were calculated for trust factor scores, RR, DET, and MaxL for purposes of direct and accurate comparisons. Finally, multiple stepwise regression analyses (Akaike information criteria, AIC) were conducted to test our hypotheses regarding the interaction-trust-failure recovery

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

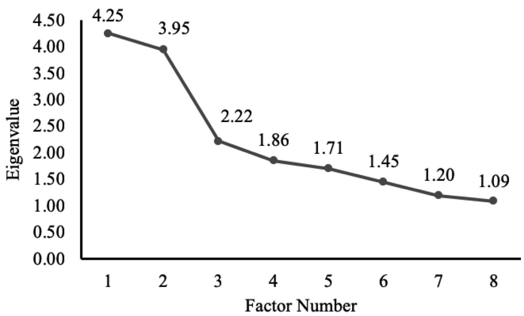

Fig. 1. Scree plot of exploratory factor analysis on 25-item.

relationships. Multiple observations from individuals and teams were assumed to be independent. JRQA was carried out in R version 3.2.3 [32] using the "crqa" [33].

A. Exploratory Factor Analysis on Trust

To reduce the dimensionality of the 25-item scale into an appropriate number of factors, an EFA based on principal axis factoring with varimax rotation was applied to the 25 trust items that were collected after the first and second sessions. To achieve an adequate sample size for the EFA, we used responses from both sessions in a single analysis [34]. According to the EFA findings, 58% of the total variance was accounted for by eight factors. The scree plot, which shows the Eigen-values for each factor (see Fig. 1), indicates a notable drop after the second factor, which explains most of the variance (the rest is "scree") [35]. Therefore, we retained the first two factors, which accounted for 17.0% and 15.8% of the variance.

As shown in Table II, Factor 1 represents the *trust that human team members put in the AI pilot role* (i.e., the mutual trust of the autonomous agent). On the other hand, Factor 2 represents *human team members' willingness to be vulnerable* (an aspect of trust [7]) with each other stemming from the trust placed in their teammates [13].

B. Trust Across the Roles and Sessions

To address the research questions, we first conducted a univariate Analysis of Variance (ANOVA) to examine how the trust that human team members put in the autonomous pilot role (Factor 1: $M = 0.00$, $SD = 0.94$) differed across the two sessions and between the navigator and photographer. Then, we conducted a separate univariate ANOVA to examine how willingness to be vulnerable (Factor 2: $M = 0.00$, $SD = 0.95$) differed across the two sessions and between the two roles. According to Levene's test, the assumption of homogeneity of variances was not violated for either of the dependent variables (for Factor 1, $F(3, 80) = 1.01, p = 0.40$; for Factor 2, $F(3, 80) = 0.72, p = 0.54$). The session's main effects on Factor 1 and Factor 2 showed evidence of statistical reliability (moderate, $p = 0.033$, and weak, $p = 0.887$, evidence, respectively; Table III).

According to Fig. 2, we interpret these results to indicate that, on average, Factor 1 decreased from Sessions 1 to 2 for both human team members. Because there was no significant

TABLE II
FACTOR LOADINGS ON THE FIRST TWO FACTORS

Factor	Item	Factor Loading
Trust that human team members put in the pilot role	I trusted the *pilot*.	0.859
	I felt the *pilot* was reliable.	0.832
	I enjoyed working with the *pilot*.	0.781
	If someone questioned the *pilot's* motives, I would give the *pilot* the benefit of the doubt.	0.514
	While chatting with the *pilot*, it felt like I was talking to a human.	0.447
	I would be comfortable giving the *pilot* a task or problem, which was critical to me, even if I could not monitor its actions.	0.494
Human team members' willingness to be vulnerable	I would tell the *pilot* about mistakes I have made on the team task, even if they could damage my reputation.	0.729
	I would share my opinion about sensitive issues with the *pilot* even if my opinion were unpopular.	0.671
	I would share my opinion about sensitive issues with the *navigator/photographer* even if my opinion were unpopular.	0.646
	I would tell the *navigator/photographer* about mistakes I have made on the team task, even if they could damage my reputation.	0.637
	If the *navigator/photographer* asked why a problem happened, I would speak freely even if I were partly to blame.	0.618
	If the *pilot* asked why a problem happened, I would speak freely even if I were partly to blame.	0.566

TABLE III
ANOVA RESULTS FOR TRUST FACTORS

Outcome	Source	*df*	*F*	*p*	η^2
Factor 1: the trust that human team members put in the AI pilot role	Session	1	4.698	0.033	0.055
	Role	1	1.875	0.175	0.023
	Session *X* Role	1	0.144	0.705	0.002
Factor 2: human team members' willingness to be vulnerable	Session	1	0.020	0.887	0.000
	Role	1	2.828	0.097	0.034
	Session *X* Role	1	1.633	0.205	0.020

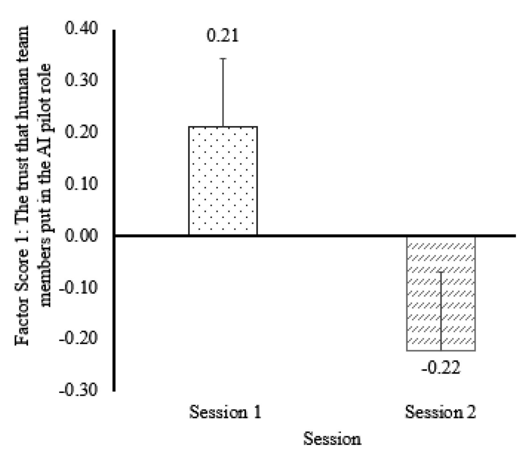

Fig. 2. Change in Factor Score 1 (trust that human team members put in the AI pilot) from Session 1 to Session 2 (vertical lines: Standard Error +SE).

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
SUMMARY OF REGRESSION ANALYSIS FOR FACTOR 1 PREDICTING PROPORTION OF AUTONOMY FAILURE RECOVERY

	B	*SE B*	*β*	*t*	*p*
Linear	0.019	0.047	0.070	0.428	0.669
Quadratic	-0.092	0.033	-0.320	-2.788	0.007

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

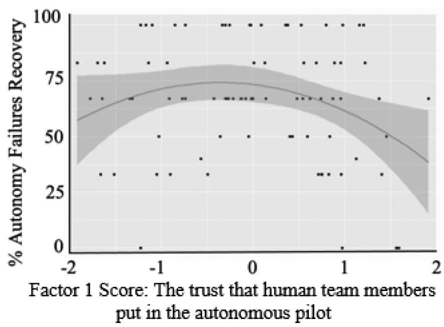

Fig. 3. Factor Score: the trust that human team members put in the autonomous pilot versus proportion of autonomy failures recovery.

difference for trust at the role level, we did not include "role" as a predictor in the follow-up analyses. Factor 2 did not differ for the navigator and photographer and did not change across sessions. The negative direction of change in Factor 1 indicates the dynamic nature of trust.

C. Trust and Failure Recovery

In order to predict autonomy and automation failure recovery across sessions using Factor 1 and Factor 2, multiple stepwise regression analyses (with AIC) were conducted (see Table IV). Although the results from the first regression indicated that the linear and quadratic terms of Factor 1 explained 12% of the variance of autonomy failures recovery, $R^2 = 0.12$, $F(4, 79) = 2.60$, $p < 0.05$, in the second regression, the only linear term of Factor 2 explained a relatively small amount of the variance in automation failure recovery, $R^2 = 0.03$, $F(1, 82) = 2.42$, $p = 0.12$.

Therefore, Factor 2, i.e., *willingness to be vulnerable,* was excluded from follow-up stepwise regression analyses because of its limited statistical evidence. Additionally, the results show that the session variable was automatically excluded from the final models. Therefore, we did not continue using the session variable in the follow-up regression analysis. In the first regression model, the quadratic term for Factor 1 had strong evidence of a statistically reliable relationship with the recovery from autonomy failures (see Fig. 3).

D. Team Interaction Dynamics

We applied Joint Recurrence Quantification Analysis (JRQA) to the multivariate team communication flow data to investigate team interaction dynamics. The basis of JRQA is the Recurrence Plot (RP [21]), which is an illustrative tool for visualizing the temporal evolution of a dynamical system when a system revisits similar states by identifying all pairs of time points in which the system returns to the same state [36]. While RPs give a visual representation of how often a single system revisits certain states or sequences of states over time, the joint recurrence RP displays the times when coupled dynamical systems visit similar states. In this article, we use discrete recurrence analysis for categorical (symbolic) time series, with symbols $0 =$ not messaging and $1 =$ messaging for each team member. Univariate discrete recurrence analysis has been used successfully to detect changing team communication dynamics in the past [37]–[39]. Due to its relative simplicity, we illustrate the concept of recurrence analysis using a univariate RP. An example RP is shown in Fig. 4.

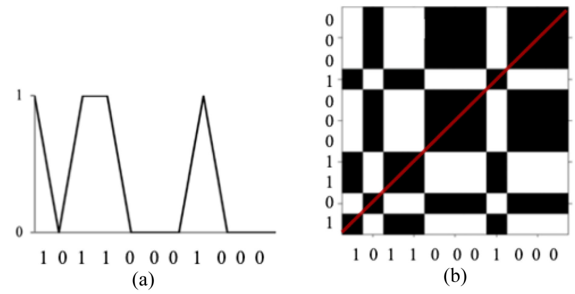

Fig. 4. (a) Example discrete time series, and (b) Discrete recurrence plot (modified from the study [23]). Recurrent "points" (block boxes) are plotted whenever "1" repeats at a later time.

Fig. 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 Fig. 4(b) is a visual representation (RP) of the time series. 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 Fig. 4(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 $x(8)$, is visually depicted in Fig. 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. Construction of JRPs follows directly from this univariate case, because the JRP is the pointwise product of all respective univariate RPs [37], i.e., RPs of each teammate. In this article, we used JRPs to examine the change in the following three commonly used JRQA measures across different window sizes: RR, DET, and MaxL. We selected a window of one minute (60-s) based on the average window size in which those three JRQA measures no longer decreased.

1) "RR" measures the overall tendency for the coupled dynamical systems (here, team members speaking) to visit the

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

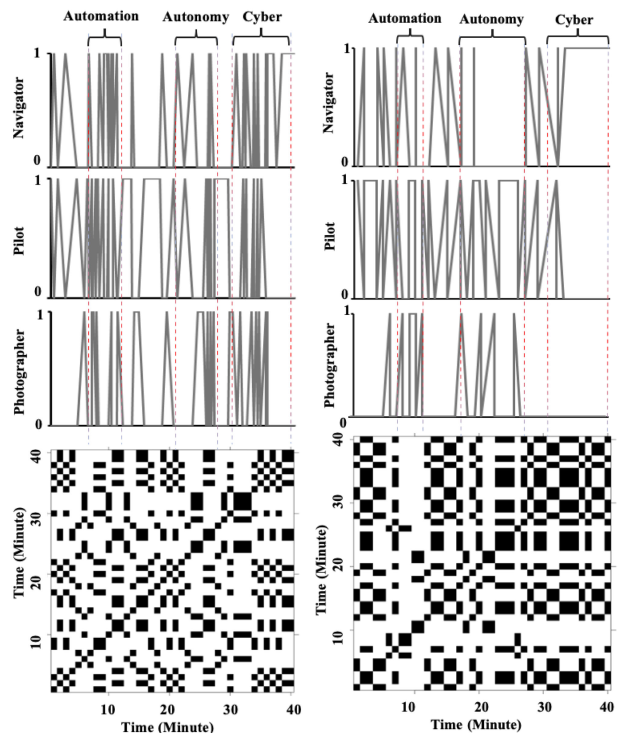

Fig. 5. Example JRPs for interactions of two RPAS teams: The high performing team recovered from all failures (left: RR = %36, DET = 56%, MaxL = 4); the low performing team only recovered from the automation failure (right: RR = 31%, DET = 51%, MaxL = 6).

same state as the proportion of recurrent points to possible recurrent points. RR is also an index of coupling strength, defined as the amount of information exchanged between individuals [40]–[42]. Here, RR corresponds to the coupling strength of team communication behavior; 2) "DET" measures the predictability of coupled systems by measuring how frequently recurrent points form repeating patterns (diagonal lines) off the main diagonal [22]. Here, DET characterizes the degree of organization of a team's communication behaviors [37], [38]: a DET of 100% indicates the input time series are perfectly repeating patterns, whereas a DET of 0% indicates the time series never repeat. In this task, DET indicates the degree to which all three team members communicated or did not communicate simultaneously in repeating patterns. We interpret DET as the predictability of team communication behavior; and 3) "MaxL" measures the stability of coupled systems by measuring the length of the longest off-diagonal sequence (pattern) of recurrent points in the RP [43]. MaxL corresponds to the stability of team communication behavior. For a binary series, it is just the longest diagonal line above the main diagonal, which is an index of attractor stability [44]–[46]. An example of two RPAS teams' JRPs for team communication flow is shown in Fig. 5, which correspond to high (see Fig 4, left) and low (see Fig. 5, right) performing RPAS teams during Mission 10 (three time series of 40-min length for each member).

The values of the three metrics for Fig. 5 plots for the high performing team are, RR = 36%, DET = 56%, MaxL = 4, and for the low performing team are, RR = 31%, DET = 51%, MaxL = 6. Each RR, DET, and MaxL was calculated based on team communication flow from each team member's message sent times, depicted in the top panel of Fig. 5. The *y*-axis shows if a message was sent by each role in a given minute: "1" for "message sent" and "0" for "no message sent." In Fig. 5 (left), the high-performing team demonstrated slightly higher coupling strength (more interaction between the team members) and more predictability in team communication behavior but was less stable than the low-performing team. According to the high coupling strength in the high-performing team, each team member communicated more frequently during each failure, and the team recovered from all of the failures. To adapt to the failures, the team also demonstrated less stable behavior than the low-performing team.

On the other hand, the low-performing team showed lower coupling strength because of a lack of interaction during the autonomy failure and cyber attack. The low-performing team members communicated more frequently during the automation failure and recovered from the failure (see Fig. 5, right). However, the same team did not demonstrate a similar communication pattern during the autonomy failure or malicious cyber attack. During the autonomy failure, the navigator did not participate, and during the malicious cyber-attack, the photographer failed to anticipate team members' needs or was unaware of the failure. In the low-performing team, higher team communication stability did not help them to recover from the failures because of a lack of team interaction.

TABLE V
SUMMARY OF REGRESSION ANALYSIS FOR INTERACTION DYNAMICS PREDICTING FACTOR 1

Term	*B*	*SE B*	*β*	*t*	*p*
Linear (*DET*)	-0.127	0.036	-1.239	-3.550	0.001
Quadratic (*DET*^{2})	0.003	0.001	0.430	2.069	0.042
Linear (*RR*)	0.169	0.054	1.142	3.141	0.002
Quadratic (*MaxL*^{2})	-0.147	0.067	-0.451	-2.194	0.031

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

E. Team Interaction Dynamics and Trust

To predict the relationship between team interaction dynamics and trust that human team members put in the autonomous pilot (i.e., Factor 1), stepwise regression analysis (AIC) was used. The results are summarized in Table V. The regression model explained 16% of the variance of human team members' trust, $R^2 = 0.16$, $\beta(4, 79) = 3.75$, $p < 0.05$. All four predictors in this model had statistically reliable associations with Factor 1: linear and quadratic terms for DET, the linear term for *RR*, and the quadratic term for MaxL, centered at their means. DET had a negative linear effect and a positive quadratic effect (after controlling for RR and maxL^2). There is a *U-shaped* relationship between DET and trust in the autonomous pilot. A moderate amount of interaction predictability was associated with the lowest levels of trust, with a minimum of trust, around 19% DET, beyond which the relationship changes from negative to positive. In other

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
SUMMARY OF REGRESSION ANALYSES FOR INTERACTION DYNAMICS PREDICTING PROPORTION OF FAILURE RECOVERY

Model	Term	*B*	*SE B*	*β*	*t*	*p*
(1) Interaction vs Autonomy	Linear (*DET*)	0.037	0.010	1.278	3.816	0.000
	Quadratic (*DET*^{2})	-0.001	0.000	-0.402	-1.914	0.059
	Linear (*RR*)	-0.077	0.015	-1.848	-5.052	0.000
	Quadratic (*MaxL*^{2})	0.004	0.001	0.968	0.001	0.000
(2) Interaction vs Automation	Linear (*DET*)	0.022	0.008	0.666	2.854	0.002
	Quadratic (*DET*^{2})	-0.001	0.000	-0.249	-1.735	0.014
	Linear (*MaxL*)	-0.120	0.053	-0.514	-2.264	0.006

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

words, increasingly predictable team interaction dynamics were negatively associated with the development of human trust in the pilot up to a point. However, there was an inflection point, beyond, which further increases in predictability was positively associated with the trust.

RR, or coupling strength, had a significant positive linear effect on trust development across the range of observed values. When human team members interact with each other and the autonomous pilot continuously during the task, the trust was also increased. MaxL, or stability, had a significant negative quadratic effect on the trust development (although the linear effect was not significant). Increasingly stable team coordination dynamics were not associated with trust put in the pilot, but an inflection point beyond which further increases stability was negatively related to the trust.

F. Team Interaction Dynamics and Failure Recovery

Model 1 from stepwise regression analysis (AIC) examines the relationship between interaction dynamics and recovery from autonomy failure (see Table VI). Overall, the model explained 28% of the variance in the proportion of autonomy failures recovery, $R^2 = 0.28$, $\beta(4, 83) = 7.91$, $p < 0.001$. The linear relationship for DET (holding constant other variables) indicates when team members had more predictable coordination, they tended to recover from autonomy failures successfully. The negative linear RR relationship indicates that decreases in coupling strength were associated with successfully recovering from autonomy failures. Finally, the positive quadratic relationship between $MaxL^2$ and autonomy failure recovery was positive, indicating that increasingly stable team interaction dynamics were positively associated with recovery from autonomy failures. One interpretation of this pattern of findings is based on the observation that the autonomous pilot did not interact effectively with the human team members during the autonomy failure. Therefore, the human team members either did not talk, and the human team members did not notice the autonomous pilot's abnormal behavior in time. This abnormal interaction during the autonomy failure was unstable than the routine conditions. The human team members demonstrated stable and persistent behavior to correct the autonomous team pilot's behavior, and an inflection point beyond which further increases stability was associated with positively related to autonomy failure recovery.

To predict the relationship between the interaction dynamics measures and automation failure recovery, stepwise regression analysis (AIC) was used (see Model 2 in Table VI). Overall, the model explained 11% of the variance in the proportion of automation failure recovery, $R^2 = 0.11$, $\beta(3, 84) = 3.28$, $p < 0.05$. The linear relationship shows a positive association between DET (holding constant other variables) and automation failure recovery, such that predictable coordination between the team members had a positive relationship with autonomy failure recovery. The significant negative quadratic DET effect replicates an earlier finding that dynamics should not be too predictable nor too random but should be in-between to adapt to novel task events [37]. Another significant finding from the model shows that the linear relationship between MaxL and automation failure recovery was negative. That is, too much stability in HAT members' interaction dynamics during automation failures was associated with a decreased ability to recover from automation failures.

V. DISCUSSION AND CONCLUSION

From the results of this article, we see two general trends. The first trend is that team interaction dynamics measures are linked to the development of different aspects of trust in HATs. This finding also verifies that team interaction dynamics measures can also be considered as objective trust measures. Specific to team interaction dynamics, aspects of coupling strength (RR), predictability (DET), and stability (MaxL) did not change from Session 1 to 2. The reason might be that Missions 2 to 10 all had failures at about the same rate, so if participants acclimated quickly, there was no change between Sessions 1 and 2 related to team interaction dynamics.

Regarding RR, the building of trust in the autonomous teammate was related to greater coupling strength. Because dynamic task environments require interaction between team members, teams need to have strong coordination through timely interaction, which will help build trust between team members to adapt to dynamic changes in the task environment. However, this interaction aspect did not help to successfully recover from the autonomy failures, because the communication between team members was not accurate due to the autonomous pilot's abnormal behavior. Being an effective team requires the exchanging of information among members in a timely manner. During the autonomy failures, the pilot either did not anticipate the human teammates' needs or demonstrated a lack of comprehension of the task.

According to the findings regarding the relationship between DET and trust, strong team communication predictability improved the human trust put in the autonomous pilot. However, moderate level predictable behavior was required for both automation and autonomy failure recovery. It makes sense because continuous and persistent interaction from the human to the autonomous teammate was also predictable. However, some amount of predictability helped teams to recover from autonomy failures. Behavioral interactions in novel conditions (during the failures) differ from routine conditions. If teams demonstrate the same predictable coordination dynamics across the task (too

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

much predictability), they could not adapt to the dynamic task environment. Therefore, teams demonstrated different types of predictable behavior, which consists of persistent behavior to the autonomous pilot, i.e., the phase transition from the routine conditions to novel conditions (i.e., autonomy and automation failures). Yet, too much predictable coordination dynamics improves trusting the autonomous team member. However, too much predictability of team coordination dynamics adversely affects team performance because of the autonomous team member's unpredictable nature in an unpredictable dynamic task environment. This opposite relationship between the team interaction predictability with human mutual trust in an autonomous team member and overcoming novel situations is clear evidence of human trust calibration needed for adaptation to the dynamic task environment.

Similarly, the automation failure recovery also required some amount of predictable coordination between the team members. That moderate predictability helped because the autonomous agent was also interacting with the human team members for automation failure recovery. This finding is logical because the autonomous pilot demonstrated unpredictable behaviors ("autonomy failures") during team performance. Future work should explore whether human team members expect more predictable behavior to build trust in the autonomous pilot or not. This finding also confirmed the previous findings from the same task [37] and human-robot teaming in urban search and rescue task [47], which underlined the inverted U-Shape model: a moderate level predictable behavior (neither too much nor too little) is required to adapt to the novel events.

According to the MaxL findings, excessive stability between team members was not beneficial in regard to increasing the trust that human team members put in the autonomous teammate, but it helped in recovering from the autonomy failures. One of the reasons for this might be that the teams demonstrated more stable communication behavior when the human team members continuously and persistently interacted with the autonomous teammate in order to recover from its abnormal behavior. In this experiment, this was the solution for autonomy failure recovery [29]. Therefore, when human team members demonstrated stability by means of continuous and persistent behavior, the team successfully recovered from autonomy failure. At the same time, they lost their trust in the autonomous pilot due to the autonomous pilot's abnormal behavior. Perturbations and team interaction dynamic measures showed us the dynamic characteristics of trust calibration. For future research in the HAT context, trust calibration characteristics need to be examined by other dynamical systems methods and measures. For instance, layered dynamics [48] is one way to examine how dynamic characteristics of trust calibration evolve based on the interaction of all the system's cognitive and technological aspects.

The second trend based on this article's findings underlines that trust and team interaction are related to autonomy and automation failure recovery. In reviewing our research questions and related hypotheses regarding how trust in HATs evolves over time, it is clear that the *trust that human team members put in the autonomous pilot* decreases as time progresses and teams continue to encounter technology failures (from Session I to Session II in our study). This finding was expected due to the nature of the overall experimental design concerning the incremental introduction of various failures.

It is possible that loss of trust over time was related to autonomy failures and not automation failures. As humans were introduced to autonomy failures caused by the autonomous pilot, their level of trust decreased. If either a human or autonomous pilot performs poorly over time in a teamwork context, a human teammate will lose trust in either entity. This reflects appropriate trust calibration over time, whether team members are human or autonomous. That being said, the exact nature of how this loss of trust occurs/develops may be different due to numerous factors (e.g., social, psychological). From this article, we can only conclude that there is a loss of trust that develops over time, which has also been seen in human-human teaming studies. Additional research is needed to compare/contrast the specific manner in how trust is lost between both types of teams. Digging deeper into the issue of loss of trust over time and interaction, we see that the development of too much or too little trust is dependent on the performance of the autonomous pilot in relation to autonomy failures. More specifically, if autonomous agent failures increase over time, the human team member naturally places less trust in the agent (again, this is similar to all-human trust dynamics), but the situation would be more complicated if the human initially overtrusted the autonomous agent. As indicated in Table IV, moderate trust was positively related to autonomy failure recovery. Previous research has brought to light the necessity that human team members understand an autonomous agent's strengths and limitations in order to maintain appropriate levels of trust, and that an agent's past performance is one of the strongest predictors of the trust that human team members will put in the autonomous agent [49]. This finding supports much of the work in the area of trust calibration and repair that indicates it is important for humans to have an appropriately calibrated amount of trust in the autonomous agent, neither too much nor too little [6].

In conclusion, team-level trust in HATs is tied to team interaction, and both are tied to failure recovery. Based on the current results, we think it is a dynamic, interaction-based process that should be measured not only through surveys but also real-time sensor-based, behavior-based, and, in general, interaction-based methods. In the future, trust development mechanisms in HATs can be envisioned that capitalize on real-time interactions between the system components.

REFERENCES

[1] N. J. McNeese, M. Demir, N. J. Cooke, and C. Myers, "Teaming with a synthetic teammate: Insights into human-autonomy teaming," *Hum. Factors*, vol. 60, no. 2, pp. 262–273, 2018, doi: 10.1177/0018720817743223.

[2] M. R. Endsley and D. B. Kaber, "Level of automation effects on performance, situation awareness and workload in a dynamic control task," *Ergonomics*, vol. 42, no. 3, pp. 462–492, Mar. 1999, doi: 10.1080/001401399185595.

[3] J. Y. C. Chen and M. J. Barnes, "Human–Agent teaming for multirobot control: A review of human factors issues," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014.

[4] J. S. Albus, "Outline for a theory of intelligence," *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, pp. 473–509, May 1991.

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.

[5] A. Ceschi, R. Sartori, and S. J. Guastello, "Advanced modeling methods for studying individual differences and dynamics in organizations: Introduction to the special issue," *Nonlinear Dyn. Psychol. Life Sci.*, vol. 22, no. 1, pp. 1–13, Jan. 2018.

[6] E. J. de de Visser, R. Pak, and T. H. Shaw, "From 'automation' to 'autonomy': The importance of trust repair in human-machine interaction," *Ergonomics*, vol. 61, pp. 1–33, Mar. 2018, doi: 10.1080/00140139.2018.1457725.

[7] D. J. McAllister, "Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations," *Acad. Manage. J.*, vol. 38, no. 1, pp. 24–59, Feb. 1995, doi: 10.2307/256727.

[8] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, Mar. 2004, doi: 10.1518/hfes.46.1.50_30392.

[9] L. Huang *et al.*, "Chapter 13 - Distributed dynamic team trust in human, artificial intelligence, and robot teaming," in *Trust In Human-Robot Interaction*," C. S. Nam and J. B. Lyons, Eds. Cambridge, MA, USA: Academic, 2021, pp. 301–319, doi: 10.1016/B978-0-12-819472-0.00013-7.

[10] D. G. Pruitt and J. Z. Rubin, *Social Conflict: Escalation, Stalemate, and Settlement*. New York, NY, USA: Random House, 1986.

[11] E. Salas, D. E. Sims, and C. S. Burke, "Is there a 'Big five' in teamwork?," *Small Group Res.*, vol. 36, no. 5, pp. 555–599, 2005, doi: 10.1177/1046496405277134.

[12] M. Deutsch, "The effect of motivational orientation upon trust and suspicion," *Human Relations*, vol. 13, pp. 123–139, 1960, doi: 10.1177/001872676001300202.

[13] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *Acad. Manage. Rev.*, vol. 20, no. 3, pp. 709–734, 1995, doi: 10.2307/258792.

[14] K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human Factors*, vol. 58, no. 3, pp. 377–400, May 2016, doi: 10.1177/0018720816634228.

[15] E. J. de Visser *et al.*, "Towards a theory of longitudinal trust calibration in human–Robot teams," *Int. J. Soc. Robot.*, Nov. 2019, doi: 10.1007/s12369-019-00596-x.

[16] C. Fletcher, "Performance appraisal and management: The developing research agenda," *J. Occupational Org. Psychol.*, vol. 74, no. 4, pp. 473–487, 2001, doi: 10.1348/096317901167488.

[17] P. Mitkidis, J. J. McGraw, A. Roepstorff, and S. Wallot, "Building trust: Heart rate synchrony and arousal during joint action increased by public goods game," *Physiol. Behav.*, vol. 149, pp. 101–106, Oct. 2015, doi: 10.1016/j.physbeh.2015.05.033.

[18] M. T. Tolston *et al.*, "Have a heart: Predictability of trust in an autonomous agent teammate through team-level measures of heart rate synchrony and arousal," in *Proc. Human Factors Ergon. Soc. Annu. Meeting*, 2018, pp. 714–715, doi: 10.1177/1541931218621162.

[19] D. Grimm, M. Demir, J. C. Gorman, and N. J. Cooke, "The complex dynamics of team situation awareness in human-autonomy teaming," in *Proc. IEEE Conf. CogSIMA*, 2018, pp. 103–109, doi: 10.1109/COGSIMA.2018.8423990.

[20] S. L. Jarvenpaa and D. E. Leidner, "Communication and trust in global virtual teams," *Org. Sci.*, vol. 10, no. 6, pp. 791–815, Dec. 1999, doi: 10.1287/orsc.10.6.791.

[21] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *EPL Europhysics Lett.*, vol. 4, no. 9, 1987, Art. no. 973, doi: 10.1209/0295-5075/4/9/004.

[22] C. L. Webber and J. P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *J. Appl. Physiol.*, vol. 76, no. 2, pp. 965–973, Feb. 1994.

[23] E. D. de Visser, R. Parasuraman, A. Freedy, E. Freedy, and G. Weltman, "A comprehensive methodology for assessing human-robot team performance for use in training and simulation," in *Proc. Human Factors Ergon. Soc. Annu. Meeting*, 2006, doi: 10.1177/154193120605002507.

[24] E. de Visser and R. Parasuraman, "Adaptive aiding of human-robot teaming: Effects of imperfect automation on performance, trust, and workload," *J. Cogn. Eng. Decis. Making*, vol. 5, no. 2, pp. 209–231, Jun. 2011, doi: 10.1177/1555343411410160.

[25] N. J. Cooke and S. M. Shope, "Designing a synthetic task environment," in *Scaled Worlds: Development, Validation, and Application*, L. R. E. Schiflett, E. Salas, and M. D. Coovert, Eds. Surrey, U.K.: Ashgate, 2004, pp. 263–278.

[26] J. Gorman, P. Amazeen, and N. Cooke, "Team coordination dynamics," *Nonlinear Dyn. Psych. Life Sci.*, vol. 14, pp. 265–289, 2010.

[27] L. D. Riek, "Wizard of oz studies in HRI: A systematic review and new reporting guidelines," *J. Human-Robot Interact*, vol. 1, no. 1, pp. 119–136, Jul. 2012, doi: 10.5898/JHRI.1.1.Riek.

[28] M. Demir, N. J. McNeese, N. J. Cooke, J. T. Ball, C. Myers, and M. Freiman, "Synthetic teammate communication and coordination with humans," in *Proc HFES Annu. Meeting*, 2015, pp. 951–955, doi: 10.1177/1541931215591275.

[29] M. Demir, N. J. McNeese, and N. J. Cooke, "The evolution of human-autonomy teams in remotely piloted aircraft systems operations," *Frontiers Commun.*, vol. 4, 2019, doi: 10.3389/fcomm.2019.00050.

[30] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task load index): Results of empirical and theoretical research," in *Human Mental Workload*, P. A. Hancock and N. Mashkati, Eds. Amsterdam, The Netherlands: North Holland, 1988, pp. 139–183.

[31] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *Int. J. Cogn. Ergon.*, vol. 4, 2000, doi: 10.1207/S15327566IJCE0401_04.

[32] R Development Core Team, "R: A programming environment for data analysis and graphics (Version 3.3.2)," Oct. 2016. Accessed: Jan. 24, 2017. [Online]. Available: https://www.r-project.org/

[33] M. I. Coco and J. Nash, "CRQA: Cross-recurrence quantification analysis for categorical and continuous time-series," 2015. Accessed: Oct. 2, 2016. [Online]. Available: https://cran.r-project.org/web/packages/crqa/index.html

[34] L. S. Tabachnick and B. G. Fidell, *Using Multivariate Statistics*. Boston, MA, USA: Allyn & Bacon, 2006.

[35] R. B. Cattell and S. Vogelmann, "A comprehensive trial of the scree and kg criteria for determining the number of factors," *Multivariate Behav. Res.*, vol. 12, 1977, doi: 10.1207/s15327906mbr1203_2.

[36] M. I. Coco and R. Dale, "Cross-recurrence quantification analysis of categorical and continuous time series: An r package," *Frontiers Psychol.*, vol. 5, Jun. 2014, doi: 10.3389/fpsyg.2014.00510.

[37] M. Demir, A. D. Likens, N. J. Cooke, P. G. Amazeen, and N. J. McNeese, "Team coordination and effectiveness in human-autonomy teaming," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 2, pp. 150–159, Apr. 2019, doi: 10.1109/THMS.2018.2877482.

[38] J. C. Gorman, D. A. Grimm, R. H. Stevens, T. Galloway, A. M. Willemsen-Dunlap, and D. J. Halpin, "Measuring real-time team cognition during team training," *Human Factors*, 2019, doi: 10.1177/0018720819852791.

[39] J. C. Gorman, N. J. Cooke, P. G. Amazeen, and S. Fouse, "Measuring patterns in team interaction sequences using a discrete recurrence approach," *Human Factors J. Human Factors Ergonom. Soc.*, vol. 54, no. 4, pp. 503–517, Aug. 2012, doi: 10.1177/0018720811426140.

[40] C. Bandt *et al.*, "Analysis of bivariate coupling by means of recurrence," in *Mathematical Methods in Signal Processing and Digital Image Analysis*, R. Dahlhaus, J. Kurths, P. Maass, and J. Timmer, Eds. Berlin, Germany: Springer, 2008, doi: 10.1007/978-3-540-75632-3_5.

[41] J. C. Gorman, T. A. Dunbar, D. Grimm, and C. L. Gipson, "Understanding and modeling teams as dynamical systems," *Frontiers Psychol.*, vol. 8, Jul. 2017, doi: 10.3389/fpsyg.2017.01053.

[42] A. J. Strang, G. J. Funke, S. M. Russell, A. W. Dukes, and M. S. Middendorf, "Physio-behavioral coupling in a cooperative team task: Contributors and relations," *J. Exp. Psychol. Human Perception Perform.*, vol. 40, no. 1, pp. 145–158, Feb. 2014, doi: 10.1037/a0033125.

[43] N. Marwan, M. Carmen Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," *Phys. Rep.*, vol. 438, no. 5/6, pp. 237–329, Jan. 2007, doi: 10.1016/j.physrep.2006.11.001.

[44] H. Abarbanel, *Analysis of Observed Chaotic Data*, 1st ed. Berlin, Germany: Springer, 1996.

[45] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis, 1 Edition*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[46] N. Stergiou, *Nonlinear Analysis for Human Movement Variability*. Boca Raton, FL, USA: CRC Press, 2016.

[47] M. Demir, N. J. McNeese, and N. J. Cooke, "Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition," *Int. J. Human-Comput. Stud.*, vol. 140, Aug. 2020, Art. no. 102436, doi: 10.1016/j.ijhcs.2020.102436.

[48] J. C. Gorman, M. Demir, N. J. Cooke, and D. A. Grimm, "Evaluating sociotechnical dynamics in a simulated remotely-piloted aircraft system: A layered dynamics approach," *Ergonomics*, vol. 62, no. 5, pp. 629–643, 2019, doi: 10.1080/00140139.2018.1557750.

[49] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, Oct. 2011, doi: 10.1177/0018720811417254.

Authorized licensed use limited to: ASU Library. Downloaded on July 26,2022 at 23:02:03 UTC from IEEE Xplore. Restrictions apply.