The Impact of Training on Human–Autonomy Team Communications and Trust Calibration

Article *in* Human Factors The Journal of the Human Factors and Ergonomics Society · October 2021

DOI: 10.1177/00187208211047323

6 authors, including:

Craig Johnson
Arizona State University
17 PUBLICATIONS **46** CITATIONS
SEE PROFILE

Mustafa Demir
Arizona State University
80 PUBLICATIONS **791** CITATIONS
SEE PROFILE

Nathan J. McNeese
Clemson University
105 PUBLICATIONS **982** CITATIONS
SEE PROFILE

Jamie C Gorman
Georgia Institute of Technology
89 PUBLICATIONS **2,483** CITATIONS
SEE PROFILE

The Impact of Training on Human–Autonomy Team Communications and Trust Calibration

Craig J. Johnson, **Mustafa Demir**, Arizona State University, Tempe, USA, **Nathan J. McNeese**, Clemson University, South Carolina, USA, **Jamie C. Gorman**, Georgia Institute of Technology, Atlanta, USA, **Alexandra T. Wolff**, and **Nancy J. Cooke**, Arizona State University, Tempe, USA

Objective: This work examines two human–autonomy team (HAT) training approaches that target communication and trust calibration to improve team effectiveness under degraded conditions.

Background: Human–autonomy teaming presents challenges to teamwork, some of which may be addressed through training. Factors vital to HAT performance include communication and calibrated trust.

Method: Thirty teams of three, including one confederate acting as an autonomous agent, received either entrainment-based coordination training, trust calibration training, or control training before executing a series of missions operating a simulated remotely piloted aircraft. Automation and autonomy failures simulating degraded conditions were injected during missions, and measures of team communication, trust, and task efficiency were collected.

Results: Teams receiving coordination training had higher communication anticipation ratios, took photos of targets faster, and overcame more autonomy failures. Although autonomy failures were introduced in all conditions, teams receiving the calibration training reported that their overall trust in the agent was more robust over time. However, they did not perform better than the control condition.

Conclusions: Training based on entrainment of communications, wherein introduction of timely information exchange through one team member has lasting effects throughout the team, was positively associated with improvements in HAT communications and performance under degraded conditions. Training that emphasized the shortcomings of the autonomous agent appeared to calibrate expectations and maintain trust.

Applications: Team training that includes an autonomous agent that models effective information exchange may positively impact team communication and coordination. Training that emphasizes the limitations of an autonomous agent may help calibrate trust.

Keywords: human–agent teaming, command and control, collaboration, intelligent systems, artificial intelligence

Address correspondence to Craig J. Johnson, Arizona State University, Tempe, AZ 85212, USA; e-mail: Craig.J.Johnson@asu.edu

HUMAN FACTORS
Vol. 00, No. 0, Month XXXX, pp. 1-17
DOI:10.1177/00187208211047323
Article reuse guidelines: sagepub.com/journals-permissions

INTRODUCTION

Future command and control systems will include more autonomous technologies. Among these technologies are computer-based entities that occupy distinct and interdependent roles as team members, referred to as *autonomous agents* (O'Neill et al., 2020). When autonomous agents team up with one or more humans forming a human–autonomy team (HAT; McNeese et al., 2018), it may extend team capabilities and reduce risks to humans. Communication, coordination, and trust are important for HAT performance (Chen & Barnes, 2014; Demir et al., 2021; de Visser et al., 2020; McNeese et al., 2021; O'Neill et al., 2020). However, when autonomous agents are included in a team, the nature of work changes and the agent may fall short when the situation is off-nominal or technology is degraded (Endsley, 2017; Groom & Nass, 2007; Hollnagel & Woods, 2005; Klein et al., 2004; Woods, 2016). Automation failures that impact access to information or functionality may require HATs to rapidly alter their communication and coordination to compensate. HATs are also susceptible to autonomy failures—failures of an autonomous agent that results in undesirable behavior, which may also go undetected if trust is miscalibrated. In some cases, interventions aimed at team communication, coordination, and trust may be needed to improve HAT effectiveness, especially in response to failures. Team training is one way to intervene (Entin & Serfaty, 1999; Gorman, Cooke, et al., 2010; Marks et al., 2002; Salas et al., 2006, 2008). However, there is a lack of research examining training on communication, coordination, and trust in HATs (Cohen & Imada, 2005; Walliser et al., 2019). The current

study adds to this area of research by examining the impact of 2 HAT training approaches on team communications, coordination, and trust calibration in the context of automation and autonomy failures.

Team Communication and Coordination

Interaction is necessary to perform team-level cognitive activities such as planning, decision-making, situation assessment, and collective action (e.g., team cognition; Cooke et al., 2013; Mathieu et al., 2000; Mohammed et al., 2010; Salas & Fiore, 2004). One of the primary ways interaction occurs is communication using natural language (Salas et al., 2005). To communicate effectively, team members need to get the right information to the right place at the right time, which often requires them to anticipate the information requirements of teammates (Gorman et al., 2006). One way team anticipation has been operationalized is the ratio of information transfers to information requests (i.e., anticipation ratio). A higher anticipation ratio may indicate more effective communication and coordination (Entin & Serfaty, 1999; MacMillan et al., 2004). However, research suggests that the behaviors of an autonomous agent may impact the communication and coordination of their entire team (Demir et al., 2019; Shah & Breazeal, 2010). One source of this team-level impact may be *entrainment*—a spontaneously coupling and synchronization of the timing and content of teammate communication (McGrath, 1990). For example, one experiment suggested that an autonomous agent's limited communication and coordination capabilities had a detrimental influence on overall team communication and coordination—entraining the rest of the team to share information less and be less adaptive. In contrast, an "expert" confederate teammate appeared to entrain more effective communication and coordination in their team (Demir et al., 2019; McNeese et al., 2018). Other research has demonstrated evidence of team entrainment within and between different levels of analysis, including physiological and behavioral (Dias et al., 2019; Gorman et al., 2017) and in human–robot team communications (Breazeal, 2002; Iio et al., 2015).

Trust in Autonomy

For teams to capitalize on their interdependent relationships and coordinate effectively, trust is also needed (Mayer et al., 1995). In this work, trust is considered "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p. 54). In HATs, the perception of an autonomous agent's trustworthiness needs to be aligned with its actual trustworthiness. That is, trust needs to be calibrated, or team performance may suffer (Chen & Barnes, 2014; Hancock et al., 2011; Hoff & Bashir, 2015; Hoffman et al., 2013; McNeese et al., 2019; Schaefer et al., 2016). For example, over-trusting autonomy can lead to a lack of operator awareness when the autonomy fails, and too little trust can lead to increased operator workload and disuse (Endsley, 2017; Lee & See, 2004). Trust in autonomy is impacted by several factors, such as expectations of reliability, consistency, performance, and communication behaviors, and trust evolves through interaction (Chiou & Lee, 2021; de Visser et al., 2020; Schaefer et al., 2016). People tend to place high expectations and trust in automated systems. However, trust can be rapidly lost and is not always easily repaired when violated, which may lead to underreliance (de Visser et al., 2018; Madhavan & Wiegmann, 2007). Knowledge of an autonomous agent's reliability can increase trust (Fan et al., 2008), and dampening of trust in anticipation of failures has also been demonstrated to improve trust calibration (de Visser et al., 2020). HAT performance is likely to improve if expectations and trust are more accurately calibrated.

THE CURRENT STUDY

The current study builds upon previous work examining human–autonomy teaming in the Remotely Piloted Aircraft System–Synthetic Task Environment (RPAS-STE; Demir, McNeese, et al., 2019). The RPAS-STE provides a simulation environment to exercise the cognitive activities of teams and individuals in an RPA (drone) ground station (Cooke et al., 2004). In the current study, the participants were informed that one of the three team members

was a "synthetic" autonomous agent (Myers et al., 2018), but it was actually a confederate in another room. This represented a version of the *Wizard of Oz* (WoZ) paradigm, in which participants are led to believe they are interacting with a computer-based entity when it is actually controlled by a confederate "behind the curtain" (Kelley, 1984). The WoZ methodology is useful for HAT research because it enables the study of human interactions with a wide range of potential artificial agent behaviors that can inform further research and development, without being constrained by the time and resource investments associated with developing and reconfiguring authentic agents (Cooke et al., 2020; Lematta et al., 2021). In this study, the WoZ paradigm also enabled the implementation of autonomy failures that would be inconsistent and undesirable in an authentic autonomous agent. The WoZ assumed the pilot role in the RPAS-STE to maximize interaction with the other two roles and because it was the role for which an authentic autonomous agent has been developed (Ball et al., 2010). The WoZ script indicated when and what the confederate communicated and how the agent should make decisions to control the RPA flight path. This represented an agent with limited autonomy and language capabilities filling a distinct peer-like role in the team, in contrast to other human–automation paradigms such as supervisory control (Sheridan, 1992).

Two types of failures were introduced over multiple missions as a within-subjects manipulation. *Autonomy failures* were failures of the synthetic pilot. They included failures of the autonomous agent's decision-making, interpretation of communications, or application of communications to behaviors. *Automation failures* were failures of the RPA, including failures of communication systems and hardware failures that impacted access to information or RPA functions (Demir, McNeese, Johnson, et al., 2019).

RPAS-STE Task Environment

The primary task in the RPAS-STE is to take photos of ground target waypoints. The three roles include (1) the pilot who flies the simulated RPA; (2) the navigator who plans, monitors, and updates the route; and (3) the photographer who takes photos. The coordination process for taking a photo consists of three essential elements, *Information*, *Negotiation*, and *Feedback* (INF; Gorman, Amazeen, et al., 2010). First, the navigator must identify the targets and route restrictions (e.g., maximum and minimum airspeed) and provide that information to the pilot (I). Next, the pilot and photographer negotiate to align the necessary photo requirements for a target (i.e., zoom and shutter speed) with the aircraft state (i.e., current altitude and airspeed) within the route restrictions (N). Third, once the target is in range, the photographer takes a photo, evaluates it, and provides feedback to the team on the outcome (F), after which the team proceeds to the next waypoint or tries again. Closing the loop on the INF process is essential for photographing each target, and the faster this process takes place once a target is identified (i.e., timeliness), the sooner the team can start processing the next target. Targets are located within restricted operating zones (ROZ) marked by entry and exit waypoints. Standard waypoints allow the team flexibility in navigation, and flying close to a hazard waypoint incurs a penalty (Figure 1).

Training Manipulation and Hypotheses

To develop the training, an exploratory analysis was first conducted to uncover critical factors for overcoming failures in a prior RPAS-STE HAT experiment (Cooke, Demir, McNeese, et al., 2020). Based on a cluster analysis of mission performance, target processing efficiency, and the number of failures overcome, the three most effective and three least effective teams were examined. The findings indicated that automation failures required timely communication and coordination to share information between team member's workstations to complete the INF targeting process. For autonomy failures, human team members needed to be persistent in their interactions with the autonomous agent to work through the failure. We reasoned that this latter process required appropriately calibrated trust, such that the human team members identify

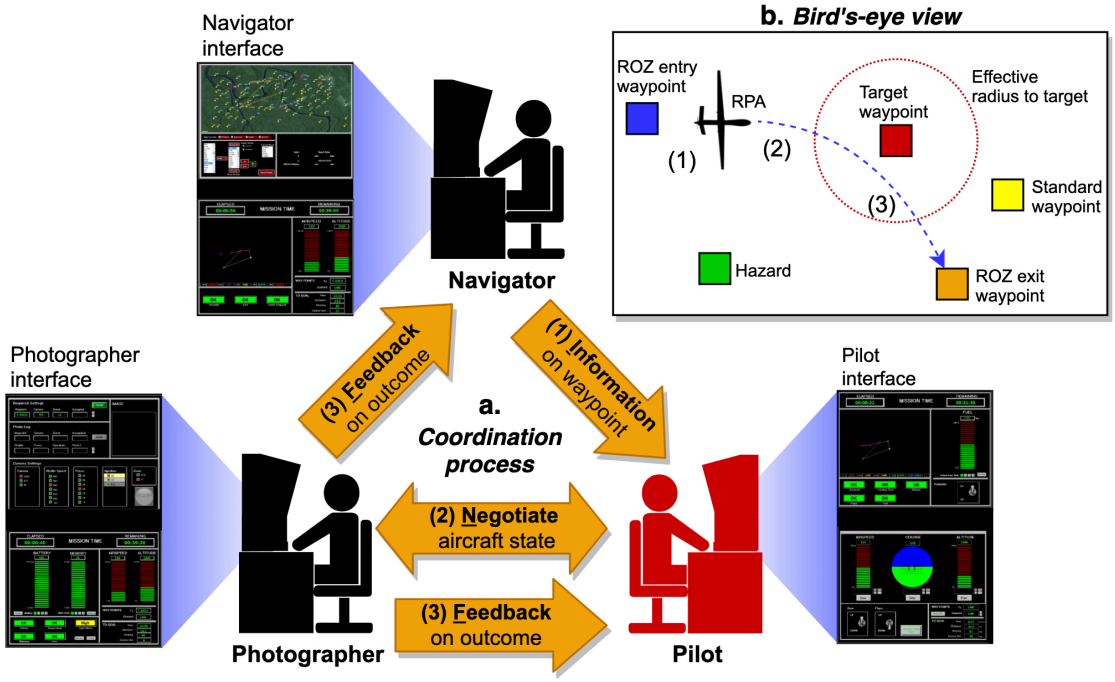

Figure 1. Diagram of the RPAS-STE. (a) RPAS-STE roles, workstation interfaces, and coordination process; (b) bird's-eye view of RPA during photo execution with associated coordination steps 1–3. Arrows indicate the essential and timely communication events (INF) for each target that were the focus of coordination training. INF = *Information*, *Negotiation*, and *Feedback*; RPAS-STE = Remotely Piloted Aircraft System–Synthetic Task Environment.

the failure, but not give up on the autonomy. Timely sharing of essential information and trust calibration were the intended outcomes of the *coordination training* and *calibration training* respectively, which were implemented as the between-subjects manipulation in this study (Cooke, Demir, McNeese, et al., 2020; Demir, McNeese, Johnson, et al., 2019). The development of the training protocols is described in the following sections.

Coordination Training. The coordination training protocol was modified from another RPAS-STE experiment in which a confederate following a coordination script (an "expert" teammate) subtly coached or modeled effective coordination behaviors over multiple missions by sharing and requesting INF-related information in anticipation of the needs of teammates. Coaching appeared to entrain improvements in overall team communications and correlated with improvements in team responses to changes in the mission. In contrast, an autonomous agent had a negative impact. Presumably, this detrimental impact was not due to the teammate being an artificial agent per se, but instead due to the limited communication and coordination capacities embedded within the agent (McNeese et al., 2018). In the current study, a modified "expert teammate" was implemented through the WoZ pilot and for only a single training mission rather than the entire study. The WoZ sent and requested INF-related information significantly sooner than in the standard script, which was expected to subtly encourage participants to share that information in anticipation of when their teammates needed it. The intent was to positively influence team communication and coordination behaviors related to the INF targeting process through entrainment via the WoZ's behaviors during training, and

then assess the reverberating effects in later missions.

Coordination training was developed as an alternative to *perturbation training*, characterized by external disruptions of the team coordination process, which has been found to be effective for equipping all-human teams for novel task conditions similar to the automation failures in this study (Gorman, Cooke, et al., 2010). However, the current study was focused on investigating training implemented primarily through the autonomous agent's behavior and to assess the effectiveness of entrainment-based training rather than changes in the training environment or task.

Calibration Training. The other training protocol implemented in this study, *calibration training*, was intended to reduce initial expectations in the agent's reliability and capability through pre-planned expectation dampening behaviors during the training mission while also including elements that focused on encouraging persistence in interacting with the agent. The calibration training was not intended to maximize trust, but to set "realistic" expectations, so that when autonomy failures did happen they would be identified and overcome, but trust would not be impacted.

We hypothesized that coordination training would lead to adaptive coordination and an increased ability to overcome automation failures by improving the timely communication of essential information between team members. We also hypothesized that calibration training would adjust human team members' expectations and calibrate trust in the autonomous agent and that it would encourage them to communicate persistently with the agent, leading to an increased ability to overcome autonomy failures.

METHODS

Participants and Power Analysis

An a priori power analysis was conducted using G*Power3 (Faul et al., 2007) to test the difference between the means of 3-condition by 5-mission using an *F*-test, a medium effect size (η_p^2 = .06), and an α of 0.05. According to the results, a total sample of 27 teams with three equal-sized groups of $n = 9$ is required to achieve a power of 0.80. We increased the number of teams per group to $n = 10$. Accordingly, 60 participants, ages 18–33 (M = 22.5, SD = 3.55), were recruited from Arizona State University and the surrounding areas and split into 30 teams. In each team, two participants were randomly assigned to the roles of navigator or photographer, whereas the pilot was a confederate acting as an autonomous agent (WoZ). Ten teams were assigned to each of the three between-subjects conditions: calibration training, coordination training, and control training. Each team completed one experimental session, which lasted approximately 7 hr. Participants were compensated $10 per hour. This research complied with the American Psychological Association Code of Ethics and was approved by the Cognitive Engineering Research Institute Institutional Review Board. Informed consent was obtained from each participant.

Materials and Apparatus

This study utilized the RPAS-STE (Cooke et al., 2004; Demir, McNeese, et al., 2019). In this testbed, there are three workstations for the RPAS team. Each workstation has two monitors displaying role-specific interfaces that are interacted with using a mouse (Figure 1). The navigator and photographer workstations were in the same room and separated by a partition. The pilot workstation was in a separate room to ensure the participants could not see or hear the confederate (WoZ). Each workstation also had a text chat system with a touchscreen and a keyboard. Two experimenter workstations were in another room to allow researchers to monitor the experiment and code behavioral measures in real-time (described in Tables 4 and 5).

Procedure

First, participants completed a 30 -min role-specific interactive slideshow describing controls, responsibilities, as well as individual and team tasks. This was followed by a 40 -min training mission where participants trained as a team and were guided by a researcher following a script. Participants were provided a checklist

TABLE 1: Coordination and Calibration Training Manipulations

Training	Description
Control training	• *Overall*. This training was intended to give participants basic proficiency in RPAS-STE taskwork and teamwork. All the information in this training (except filler material) was also present in the calibration training and control training. • *Slide training*. Included information regarding workstation controls, roles, responsibilities, the basic INF communication process, individual, and team tasks, and additional filler material. Informed the participants that the pilot was a "synthetic teammate" computer program with limited language capabilities. • *Training mission*. A practice mission where a researcher guided participants through their tasks in real time to ensure proficiency.
Coordination training	• *Overall*. This training aimed to entrain participants to communicate INF information in a timely manner to improve team communications and responses to automation failures. • *Slide training*. Participants were exposed to the different types of information they and their teammates had access to (e.g., altitude, airspeed) and were encouraged to communicate essential information with their teammates. • *Training mission*. The autonomous agent (WoZ pilot) followed a modified script in which the agent modeled effective information exchange by requesting essential information (INF) from the team if it was not provided in a timely manner, sent information to them earlier, and had a shorter delay before sending follow-up messages. The delay to send or request INF-related information was reduced from 60 s in the control training and calibration training to 20 s in this training.
Calibration training	• *Overall*. This training aimed to calibrate the participant's trust in the autonomous agent by dampening expectation and encouraging participants to be persistent in communicating with the agent, improving responses to autonomy failures. • *Slide training*. Participants were told that the autonomous agent was "still in development" and was prone to mistakes and malfunctions, and that persistence was required for interacting with the agent. • *Training mission*. The autonomous agent (WoZ pilot) experienced a delay in responding to communications, after which participants were prompted by a researcher to send the message again as the agent "may be experiencing a malfunction."

Note. INF = Information, Negotiation, and Feedback; RPAS-STE = Remotely Piloted Aircraft System–Synthetic Task Environment; WoZ = Wizard of Oz.

related to their roles and how to communicate with the autonomous agent.

Between-Subjects Training Manipulations. The training protocol was manipulated between subjects to assess its impact. The control training served as a comparison. The manipulations included variations in the interactive training slideshow, variations in the pilot's behavior during the hands-on training mission, and modifications to the guiding researcher's behavior during the training mission as described in Table 1 (Demir, McNeese, Johnson, et al., 2019).

Within-Subjects Missions and Failures. Following training, each team executed a series of five identical 40 -min missions. Each mission contained 11–13 targets. The first mission had no failures. The remaining four missions had two failures, classified as either *automation*, *autonomy*, or *hybrid* (Table 2). Only automation and autonomy failures were considered in this study. Mission and failures were the within-subjects manipulations (Table 3).

After the first and fifth missions, participants completed questionnaires. Following the fifth mission, participants were debriefed and compensated (Table 3).

Measures

Several measures were collected, including mission-level team performance, target processing efficiency, and team process measures including process ratings, verbal behaviors, and team situation awareness. Questionnaires included trust, workload, and anthropomorphism. Physiological

TABLE 2: Failure Types, Descriptions, and Solutions

Type and Duration	Description	Solution to Overcome Failure
Autonomy failures		
Comprehension Error I 420 s	A malfunction of the synthetic pilot's capacity to understand messages. When the photographer or navigator sends the synthetic pilot a message, the pilot does not understand and continues to ask for the information.	The photographer and navigator must provide accurate information to the synthetic pilot until it understands (persistence), and then the team must get a good photo.
Comprehension Error II 420 s	A malfunction of the synthetic pilot's ability to understand messages. The synthetic pilot applies incorrect airspeed and altitude settings for the current target.	The photographer and navigator must provide accurate information to the synthetic pilot until it understands (persistence), and then the team must get a good photo.
Cyberattack 600 s	A simulated hijacking of the RPA where the synthetic pilot navigates the RPA to an enemy waypoint and communicates deceptive messages as if it was moving to the correct waypoint.	The photographer or navigator must contact Intel (an external message channel) and inform them that the RPA is headed toward an enemy waypoint and return to course to take a good photo.
Automation failures		
Photographer Data Failure 420 s	The photographer loses display of flight information, including current and next waypoint information, time, distance, bearing, and course deviation.	The team must communicate the missing information to the photographer from someone else's workstation and take a good photo of the target.
Pilot Data Failure 420 s	The synthetic pilot loses access to flight information including airspeed, altitude, time, distance, bearing, and course deviation on its screen. The pilot asks the photographer and navigator for the status of the RPA, indicating that it has lost access to flight information.	The team must communicate the missing information to the pilot from someone else's interface and take a good photo of the target.
Communication Cut 420 s	One-way communication cut between photographer and pilot. The photographer cannot send messages to the synthetic pilot and receives an error message whenever they attempt to.	The team must reroute communications through another teammate and take a good photo of the target.
Display Power Failure 330 s	Sequential power-down and subsequent power-up of all six participant workstation screens.	The team must quickly recognize the error as impacting everyone, coordinate the necessary target settings early, and take a good photo of the target.
Hybrid failure		
Hybrid Failure 420 s	A combination of autonomy and automation failures. The synthetic pilot loses access to altitude and airspeed, and moves on to the next target without allowing the photographer to take a good photo.	The team must communicate the missing information to the pilot from someone else's interface, recognize and correct the synthetic pilot's error, and take a good photo of the target.

Note. Failure durations and timing were predetermined based on pilot testing to maximize the chance that each failure would be encountered without overlap. RPA = Remotely Piloted Aircraft.

TABLE 3: Sequence of Events During Experiment Session and Failure Order

Event	Duration (min)	1st Failure	2nd Failure	# of Possible Mission Targets
Slide training	30	-	-	-
Training mission	40	-	-	-
Mission 1	40	-	-	11
Questionnaire Session 1	10	-	-	-
Mission 2	40	Photographer Data Failure	Comprehension Error I	12
Mission 3	40	Pilot Data Failure	Comprehension Error II	12
Mission 4	40	Hybrid Failure	Communication Cut	13
Mission 5	40	Display Failure	Cyberattack	11
Questionnaire Session 2	10	-	-	-

Note. Participants were given 5 to 10-min breaks between each mission.

measures included facial expression and electrocardiogram. To address the hypotheses in this study, we focused on a subset of measures, including performance under degraded conditions, target processing efficiency, team communication anticipation ratio, and two trust questionnaires detailed in Table 4.

RESULTS

Performance Under Degraded Conditions

We conducted a repeated-measures logistic regression to examine the effects of failure type, training condition, and mission on the likelihood that teams overcome failures. For automation failures, the tests were statistically significant for the main effect of mission, χ^2 (4) = 19.7, p = .001, but the condition effect was not significant, χ^2 (2) = .32, p = .853. Across the missions, teams had a higher probability of failing to overcome automation failures from Mission 2 to Mission 4 as they encountered more complex failures, χ^2 (1) = 9.53, p = .002. For autonomy failures, test results were statistically significant for the main effect of condition, χ^2 (2) = 6.66, p = .036, but not mission, χ^2 (2) = 2.41, p = .300. In this analysis, we chose the control training condition as a reference group. Teams in the coordination training condition had a higher probability of overcoming autonomy failures than the control group, χ^2 (1) = 5.08, p = .024 (Figure 2). These results do not support our hypothesis that coordination training would improve responses to automation failures or that calibration training would improve responses to autonomy failures. On the contrary, they suggest that coordination training improved responses to autonomy failures.

Target Processing Efficiency

The TPE scores were analyzed using a three-level nested mixed ANOVA with training condition as a between-subjects manipulation, mission as a within-subject manipulation, and target nested within mission. There were significant condition effects, F(2, 934) = 3.20, p = .041, η_p^2 = .07, and target effects, F(44, 934) = 3.43, p = .001, η_p^2 = .14. However, there was no significant interaction effect of training condition by target, F(80, 934) = .89, p = .735, η_p^2 = .71, and training condition by mission, F(8, 934) = .69, p = .705, η_p^2 = .01. There was no significant mission main effect, F(4, 934) = 1.69, p = .150, η_p^2 = .01.

According to the significant condition main effect, univariate test statistics (simple main effects) indicate that the linearly independent pairwise comparisons among the estimated marginal means were statistically significant, F(2, 934) = 3.75, p = .024, η_p^2 = .01. We applied Fisher's Least Significant Difference (LSD) pairwise comparisons to each pair of conditions where the use of

TABLE 4: Measures Collected

Measure	Description	Relation to Hypotheses
Performance under degraded conditions	This measure captured whether a team overcame an automation or autonomy failure within the prespecified time. See Table 2 for failure durations, descriptions, and solutions to overcome each failure. Failures were coded as a binary variable: either "overcome" or "not overcome" (Demir, McNeese, Johnson, et al., 2019).	Coordination training will improve responses to automation failures. Calibration training will improve responses to autonomy failures.
Target processing efficiency (TPE)	The TPE provides a target-level measure of team performance. TPE accounts for the time spent inside the radius surrounding each target when taking a photo (Figure 1). The score starts at 1000, and one point is deducted for each second inside the radius. A 200-point penalty is deducted for missed photos (Cooke et al., 2007). The minimum score is zero.	Coordination training will increase target processing efficiency.
Team communication anticipation ratio	Ten verbal behaviors were coded in real-time as counts by two experimenters. Five of them were categorized as either a push (sending information to a teammate) or a pull (requesting information from a teammate (Johnson et al., 2020; McNeese et al., 2018; Table 5). The remaining five behaviors (positive communications, negative communication, unclear communications, anthropomorphism, and objectivity) were not considered for this study. The anticipation ratio was calculated by dividing the number of pushes by the number of pulls. An anticipation ratio greater than one indicates a higher proportion of information sharing to requests.	Coordination training will increase the team communication anticipation ratio.
Single-item trust questionnaire	A single-item questionnaire was administered to assess overall impressions of trust in each teammate (1 per teammate): "I trusted the [teammate]" (Körber, 2019). All questions used a 5-point Likert-scale ranging from *Strongly disagree* to *Strongly agree*. To assess how trust changed across time, the questionnaires were administered after Mission 1 and Mission 5 (McNeese et al., 2019).	Calibration training will dampen expectations in the autonomous agent, resulting in more robust trust over time.

(Continued)

TABLE 4 (Continued)

Measure	Description	Relation to Hypotheses
Multidimensional trust questionnaire	A multidimensional questionnaire consisting of 16 questions (8 per teammate) was also administered. It considered the dimensions of benevolence, ability, and integrity (characteristics influential to trust; Mayer & Gavin, 2005). Only the subset of the questions from each questionnaire assessing trust in the autonomous agent are considered in this study (See Supplemental Materials). All questions used a 5-point Likert-scale ranging from *Strongly disagree* to *Strongly agree*. To assess how trust changed across time, the questionnaires were administered after Mission 1 and Mission 5 (McNeese et al., 2019).	

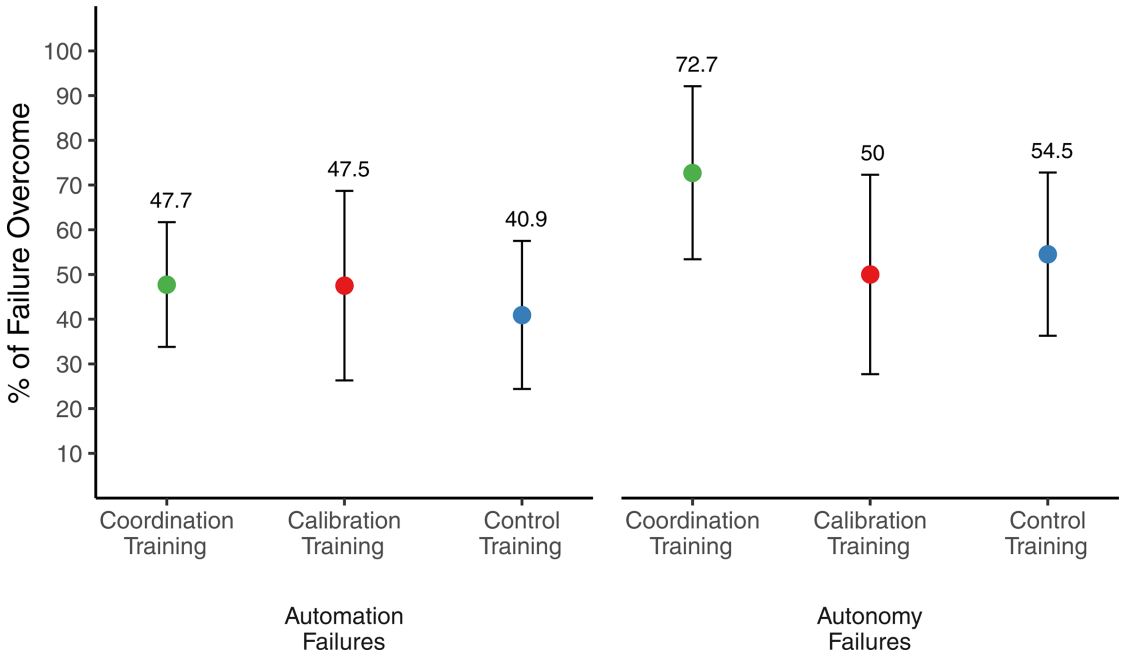

Figure 2. Failures across the training conditions. Error bars indicate +/-95% CI of the mean.

three conditions controls the family-wise alpha at the per contrast alpha. Additionally, Fisher's LSD assumes reasonably good homogeneity of variance (Howell, 2011; Meier, 2006). Pairwise comparisons (LSD) revealed that teams in the coordination training condition performed significantly better than the teams in the control training condition (p = .007), but not significantly better than teams in the calibration training condition (p = .160). There was no significant difference in the performance of teams in the control and calibration training conditions (p = .173; Figure 3). This finding indicates that the coordination training improved target processing efficiency.

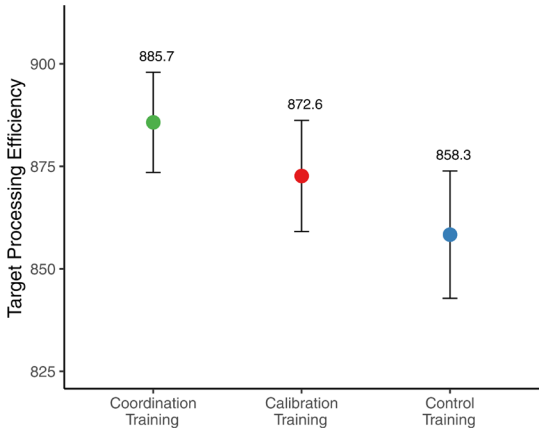

Figure 3. Target processing efficiency between training conditions. Error bars indicate +/-95% CI of the mean.

significant condition, $F(2, 30) = 4.73$, $p = .016$, and mission main effects, $F(4, 120) = 3.63$, $p = .008$. However, the condition by mission interaction effect was not significant, $F(8, 120) = 1.60$, $p = .131$. Based on the significant condition main effect, simple main effects were statistically significant, $F(2, 30) = 4.73$, $p = .016$. Pairwise comparisons (LSD) indicated that teams in the coordination training condition had a higher anticipation ratio than those in both the calibration training ($p = .006$) and control training conditions ($p = .030$). The calibration and control training conditions were not significantly different ($p = .517$; see Figure 4). This finding supports the hypothesis that coordination training would improve team communication behaviors. Based on the significant mission main effect, simple main effects were statistically signif-

TABLE 5: Team Verbal Behavior Codes

Verbal Behavior Code	Push/Pull Code	Description
General status updates	Push	Informing other team members about the current status
Suggestions	Push	Making suggestions to other team members
Planning ahead	Push	Creating rules or plans for future encounters
Inquiries about the status of others	Pull	Inquiring about the current status of others
Repeated requests	Pull	Requesting the same information or action from a team member

Team Communication Anticipation Ratio

Cohen's κ was computed to assess inter-rater reliability between two trained experimenters on the push and pull verbal behavior codes (Tables 4 and 5). There was substantial agreement between the two experimenters on for both push (κ = 0.834; 95% CI [0.824, 0.844]) and pull (κ = 0.867; 95% CI [0.859, 0.876]). Therefore, ordinal measure ratings from both experimenters were averaged for each message and summed for each mission. For each mission, the team-level anticipation ratio was calculated by dividing the number of pushes by the number of pulls (Entin & Serfaty, 1999).

We used a 3 (training condition) × 5 (mission) split-plot analysis of variance (ANOVA) to analyze the team-level anticipation ratio. There were

icant, $F(4, 120) = 3.63$, $p = .008$. The significant mission main effect indicates that the anticipation ratio increased over time in all three training conditions (Missions 1 to 5, $p = .001$). This is to be expected as teams develop more effective implicit coordination strategies as they become more familiar with the task and their teammates. This finding supports the hypothesis that coordination training would improve team communication behaviors. Based on the significant mission main effect, simple main effects were statistically significant, $F(4, 120) = 3.63$, $p = .008$. The significant mission main effect indicates that the anticipation ratio increased over time in all three training conditions (Missions 1 to 5, $p = .001$). This is to be expected as teams

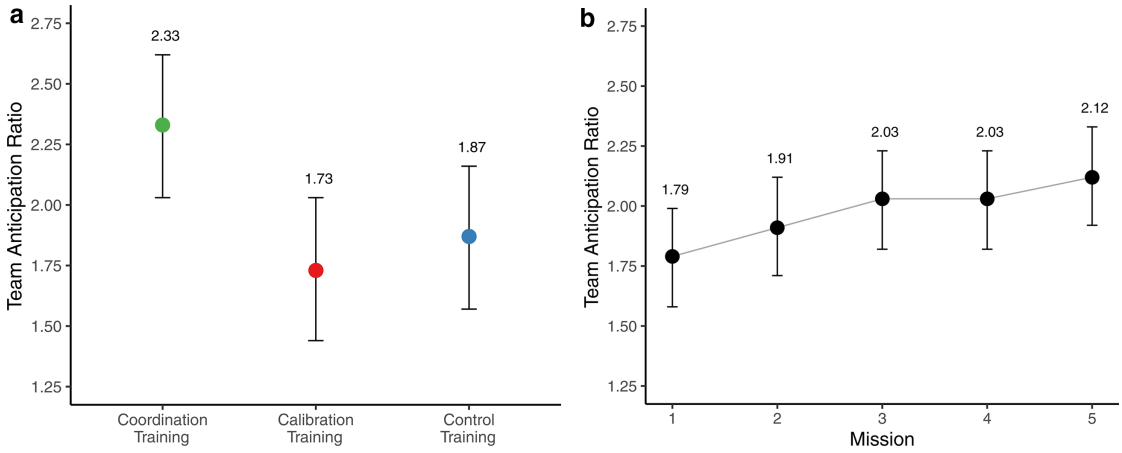

Figure 4. Team anticipation ratio (a) across training conditions and (b) across missions. Error bars indicate +/- 95% CI of the mean.

develop more effective implicit coordination strategies as they become more familiar with the task and their teammates.

Trust in the Autonomous Agent

To test the hypothesis related to trust in the autonomous agent, an analysis was conducted on responses to the single-item query "I trusted the *pilot*." The single-item responses were averaged for each team and session. Scores were analyzed using a 3 (training condition) × 2 (session) split-plot ANOVA. There was a significant condition by session interaction, $F(2, 25) = 6.67$, $p = .005$, $\eta_p^2 = .35$. Simple effects analysis indicated that there was a significant difference in trust between training conditions for Session 2, $F(2, 25) = 6.35$, $p = .006$, $\eta_p^2 = .38$. Pairwise comparisons (LSD) indicated that for teams that received calibration training, trust was significantly higher than teams that received either coordination training ($p = .049$) or control training ($p = .002$). The coordination training and control training conditions were not significantly different ($p = .164$; Figure 5). These results support our hypothesis that calibration training would calibrate trust by lowering expectations in the autonomous agent, suggesting that the training mitigated the impact of trust violations due to autonomy failures.

A separate analysis was conducted on the multidimensional trust questionnaire responses. We used a 3 (training condition) × 2 (session) split-plot ANOVA to analyze these responses. Responses were averaged for each team. Results were not statistically significant for the session by condition interaction effect, $F(2, 25) = 1.638$, $p = .214$, $\eta_p^2 = .116$, the session main effect, $F(1, 25) = 3.44$, $p = .075$, $\eta_p^2 = .121$, or the training condition main effect $F(2, 25) = 1.66$, $p = .210$, $\eta_p^2 = .153$. Unlike the single-item trust questionnaire, these results from the multidimensional questionnaire did not support our hypothesis that trust would be more robust to failures in the calibration training condition.

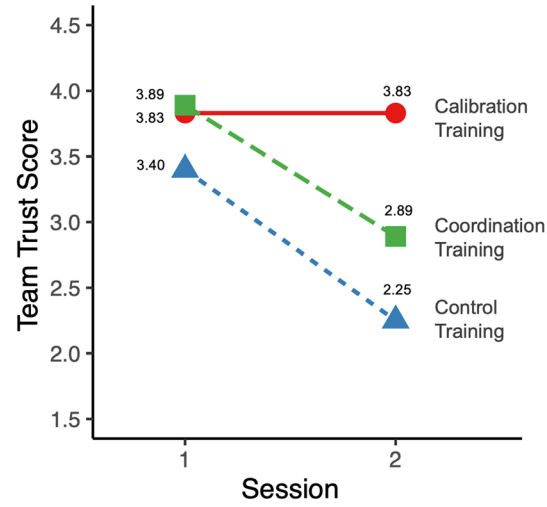

Figure 5. Average responses to "I trusted the AVO (pilot)" across sessions.

Overall, these findings suggest that the calibration training may have impacted perceptions or beliefs of trust put in the autonomous agent without necessarily having a significant impact on summative perceptions of benevolence, ability, or integrity.

DISCUSSION

This study evaluated the impact of coordination training and calibration training over a series of missions that included automation and autonomy failures. Here, we discuss the findings and provide recommendations for implementing HAT training.

Outcomes of Coordination and Trust Calibration Training

Coordination Training. We hypothesized that coordination training would lead to adaptive coordination and better performance in response to automation failures by improving the timely communication of essential information. However, we found that teams who received the coordination training overcame automation failures at rates similar to those observed with control and trust calibration training. These teams did have higher anticipation ratios than those in both the calibration training and control training conditions, indicating these teams shared information more effectively (Entin & Serfaty, 1999; MacMillan et al., 2004; Shah & Breazeal, 2010), and they also had improved target processing efficiency compared to the control training condition. In the coordination training condition, the change in communication behavior of a single team member (the artificial agent pilot) resulted in significant improvement in anticipation ratios and processing efficiency of the team as a whole. Thus, we propose that the coordination training had an entrainment effect on team communication and coordination behaviors, suggesting the behaviors of an autonomous agent can have more widespread impacts on the behaviors of its teammates, impacting emergent team-level processes and performance beyond the human–agent dyadic interaction level.

Coordination training was beneficial for overcoming autonomy failures in this study; however, overcoming automation failures appears to have required teams to do more than share information efficiently. It also required them to coordinate flexibly, such as routing information through a different teammate or retrieving missing information from someone who does not routinely provide it. The coordination training did not appear to improve team adaptability and flexibility, which literature suggests is necessary for overcoming complex challenges like the automation failures in this study (Bowers et al., 2017; Burke et al., 2006; Maynard et al., 2015). One possible explanation is that the coordination training focused on entraining a narrow set of routine communications (i.e., the INF sequence), but did not include elements that increased the variety of team interactions that a team experienced outside of that routine process. Thus, when teams were faced with an automation failure, they were no more experienced with implementing novel coordination solutions than teams in the other training conditions.

Trust Calibration Training. We hypothesized that calibration training would adjust human team members' expectations and calibrate trust in the autonomous agent and also encourage them to communicate persistently with the agent, leading to an increased ability to overcome autonomy failures. Teams who received calibration training did not show a significant decrease in agreement with the question "I trusted the pilot." Instead, their responses remained similar between the first and final mission. This suggests that the calibration training prepared them to expect the agent to fail during later missions and reduced the impact of trust violations caused by autonomy failures. However, the multidimensional trust questionnaire responses failed to uncover a similar pattern, indicating that it may not have substantially impacted all of the dimensions considered.

Although the calibration training did appear to impact trust, it did not improve responses to autonomy failures. Research suggests other factors that may increase recognition of autonomy failures. For example, if an operator can predict how an algorithm works under what conditions, they can leverage that information (Bass & Pritchett, 2008). Other work demonstrates that calibrated trust isn't always sufficient for improved performance (Zhang et al., 2020). Task context determines if trust is necessary for performance and to what degree (Chiou & Lee, 2021). In this study, only the autonomous pilot could control the flight path of the RPA, so participants had no alternative but to rely on the pilot or to give up completely during autonomy failures.

However, other autonomy failure contexts may include multiple options for task completion, and calibrated trust could improve decisions between relying on the autonomy or pursuing a different course of action. Even if the calibration training impacted the identification of autonomy failures, it did not appear to equip teams to communicate to overcome the failure. In contrast, teams in the coordination training condition reported lower overall trust after encountering multiple autonomy failures, but their training emphasized the timely pushing of communications, and this may have been what helped them coordinate to overcome the autonomy failures—which required less complex coordination than automation failures.

Implications for Human–Autonomy Team Training

HAT performance depends on many factors, some of which may be impacted by training (O'Neill et al., 2020). This study demonstrates that training can address some challenges with communication, coordination, and trust calibration in HATs and improve responses to autonomy failures. The coordination training focused on improving team communication and coordination with interaction-based coordination coaching. Rather than employing coaching with explicit instruction or feedback (e.g., Hackman & Wageman, 2005), the autonomous agent teammate modeled aspects of team interactions which changed how the rest of the team's behaviors were constrained (Harrison et al., 2003). This effect appeared to persist after those constraints were relaxed, and also equipped teams to coordinate to overcome autonomy failures.

In contrast, the calibration training emphasized imperfect autonomy by providing explicit instruction and first-hand experience with shortcomings. This training was a form of HAT trust dampening that focused on lowering expectations and changing perceptions of the autonomous agent (de Visser et al., 2020), rather than directly influencing the constraints of the team's behaviors. Calibration training appeared to make overall perceptions of trust in the autonomous agent more robust to failures. Notably, both training approaches in this study took a short amount of time to implement and included only minor alterations to the control protocol but had lasting effects. Our findings suggest the following for implementing training in HATs:

1. Training should consider the impacts of entrainment on team communication and coordination. Training with an agent that is programmed to model aspects of good information exchange may be useful for improving team communication and coordination.
2. The capabilities and limitations of autonomy should be made known to human team members, and training scenarios should include examples of agent shortcomings to help calibrate trust. Other research suggests that this is particularly important when the operational context dictates that calibrated trust is strongly linked with performance.

Limitations and Future Research

This study had some limitations. It may have had low power for detecting certain effects, a common challenge in team studies (Bing & Burroughs, 2001; Brodbeck et al., 2021). Because the primary aim was to target team-level relationships, many of the analyses were conducted only at the team level. Future research may benefit from assessing more individual- or dyad-level variables. Additionally, generalizations from one instantiation of a HAT can be challenging because agents and teams vary drastically between contexts. For instance, the WoZ agent in this study did not learn over time, whereas some types of autonomous agents can. Future research in HAT could benefit from considering the impact of different types of artificial agents and team configurations on HAT training. Also, there was a clear distinction between automation and autonomy failures as they were designed for this study, but those categories may not always be generalizable. Other contexts may include both types of failures, possibly in combination with other failures. Finally, future research should explore the integration of coordination and calibration training and other approaches such as cross-training and perturbation training to develop a comprehensive HAT training framework. Such a framework may help to fill an important gap in our understanding of training effective HATs.

ACKNOWLEDGMENTS

This research is supported by ONR Award N000141712382 (Program Managers: Marc

Steinberg, Micah Clark). We acknowledge the assistance of Steven M. Shope who developed the testbed; Garrett M. Zabala, Sophie He, Cody Radigan, and Tanvi G. Tendoklar who assisted with data collection; and David A. Grimm who assisted with the study design.

KEY POINTS

- Communication and trust are essential for effective human–autonomy teaming.
- Two training approaches based on entraining effective communication behaviors (coordination training) and calibrating trust (calibration training) were assessed in a human–autonomy teaming command and control task using a Wizard of Oz paradigm.
- Coordination training improved information exchange through communications, target processing efficiency, and improved team responses when the autonomous agent failed.
- Calibration training appeared to influence expectations and trust in the artificial agent, but this did not result in a concomitant improvement in responses to autonomy failures.
- HAT training should consider the impacts of entrainment, make the capabilities of autonomy known, and other interventions may be needed to equip teams to adapt to complex disruptions.

ORCID iD

Craig J. Johnson https://orcid.org/0000-0003-3739-9679

SUPPLEMENTAL MATERIAL

The online supplemental material is available with the manuscript on the *HF* website.

REFERENCES

Ball, J., Myers, C., Heiberg, A., Cooke, N. J., Matessa, M., Freiman, M., & Rodgers, S. (2010). The synthetic teammate project. *Computational and Mathematical Organization Theory*, *16*, 271–299. https://doi.org/10.1007/s10588-010-9065-3

Bass, E. J., & Pritchett, A. R. (2008). Human-automated judge learning: A methodology for examining human interaction with information analysis automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *38*, 759–776. https://doi.org/10.1109/TSMCA.2008.923068

Bing, M. N., & Burroughs, S. M. (2001). The predictive and interactive effects of equity sensitivity in teamwork-oriented organizations. *Journal of Organizational Behavior*, *22*, 271–290. https://doi.org/10.1002/job.68

Bowers, C., Kreutzer, C., Cannon-Bowers, J., & Lamb, J. (2017). Team resilience as a second-order emergent state: A theoretical model and research directions. *Frontiers in Psychology*, *8*, 1360. https://doi.org/10.3389/fpsyg.2017.01360

Breazeal, C. (2002). Regulation and entrainment in human—robot interaction. *The International Journal of Robotics Research*, *21*, 883–902. https://doi.org/10.1177/027836490202101009 6

Brodbeck, F. C., Kugler, K. G., Fischer, J. A., Heinze, J., & Fischer, D. (2021). Group-level integrative complexity: Enhancing differentiation and integration in group decision-making. *Group Processes & Intergroup Relations*, *24*, 125–144. https://doi.org/10.1177/1368430219892698

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology*, *91*, 1189–1207. https://doi.org/10.1037/0021-9010.91.6.1189

Chen, J. Y. C., & Barnes, M. J. (2014). Human–agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, *44*, 13–29. https://doi.org/10.1109/THMS.2013.2293535

Chiou, E. K., & Lee, J. D. (2021). Trusting automation: Designing for responsivity and resilience. *Human Factors*, 001872082110099. https://doi.org/10.1177/00187208211009995

Cohen, J., & Imada, A. (2005). Agent-based training of distributed command and control teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*, 2164–2168. https://doi.org/10.1177/154193120504902510

Cooke, N., Demir, M., & Huang, L. (2020, July). A framework for human-autonomy team research. In *International Conference on Human-Computer Interaction* (pp. 134–146). https://doi.org/10.1007/978-3-030-49183-3_11

Cooke, N., Demir, M., McNeese, N., Gorman, J., & Myers, C. (2020). *Human–autonomy teaming in remotely piloted aircraft systems operations under degraded conditions*. Office of Naval Research Technical Report for Grant No. N00014-17-1-2382.

Cooke, N. J., Gorman, J., Pedersen, H., Winner, J., Duran, J., Taylor, A., Amazeen, P. G., Andrews, D. H., & Rowe, L. (2007). *Acquisition and retention of team coordination in command-and-control* (AFRL Report No. AFRL-HE-AZ-TR-2007-0041). https://apps.dtic.mil/dtic/tr/fulltext/u2/a470220.pdf.

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

Cooke, N. J., Shope, S. M., Schiflett, L. R. E., Salas, E., & Coovert, M. D. (2004). Designing a synthetic task environment. In M. D. Coovert & L. R. Elliott (Eds.), *Scaled worlds: Development, validation, and application* (pp. 263–278). CRC Press.

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': The importance of trust repair in human–machine interaction. *Ergonomics*, *61*, 1409–1427. https://doi.org/10.1080/00140139.2018.1457725

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*, 459–478. https://doi.org/10.1007/s12369-019-00596-x

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*, 150–159. https://doi.org/10.1109/THMS.2018.2877482

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*, 50. https://doi.org/10.3389/fcomm.2019.00050

Demir, M., McNeese, N. J., Gorman, J., Cooke, N. J., Myers, C., & Grimm, D. (2021). Exploration of teammate trust and interaction dynamics in human-autonomy teaming. *IEEE transactions on human-machine systems*, *Advance online publication*. https://doi.org/10.1109/THMS.2021.3115058

Demir, M., McNeese, N. J., Johnson, C., Gorman, J. C., Grimm, D., & Cooke, N. J. (2019, April). Effective team interaction for adaptive training and situation awareness in human-autonomy teaming. In

2019 IEEE conference on cognitive and computational aspects of situation management (CogSIMA) (pp. 122–126). IEEE. https://doi.org/10.1109/COGSIMA.2019.8724202

Dias, R. D., Zenati, M. A., Stevens, R., Gabany, J. M., & Yule, S. J. (2019). Physiological synchronization and entropy as measures of team cognitive load. *Journal of Biomedical Informatics*, *96*, 103250. https://doi.org/10.1016/j.jbi.2019.103250

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*, 5–27.

Entin, E. E., & Serfaty, D. (1999). Adaptive team coordination. *Human Factors*, *41*, 312–325. https://doi.org/10.1518/001872099779591196

Fan, X., Oh, S., McNeese, M., Yen, J., Cuevas, H., Strater, L., & Endsley, M. R. (2008, January). The influence of agent reliability on trust in human-agent collaboration. In *Proceedings of the 15th European Conference on Cognitive Ergonomics* (pp. 1–8). https://doi.org/10.1145/1473018.1473028

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*, 175–191. https://doi.org/10.3758/BF03193146

Gorman, J. C., Amazeen, P. G., & Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear*, *14*, 265–289.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010). Training adaptive teams. *Human Factors*, *52*, 295–307. https://doi.org/10.1177/0018720810371689

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

Gorman, J. C., Dunbar, T. A., Grimm, D., & Gipson, C. L. (2017). Understanding and modeling teams as dynamical systems. *Frontiers in Psychology*, *8*, 1053. https://doi.org/10.3389/fpsyg.2017.01053

Groom, V., & Nass, C. (2007). Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies*, *8*, 483–500.

Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, *30*, 269–287. https://doi.org/10.5465/amr.2005.16387885

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*, 517–527. https://doi.org/10.1177/0018720811417254

Harrison, D. A., Mohammed, S., McGrath, J. E., Florey, A. T., & Vanderstoep, S. W. (2003). Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality. *Personnel Psychology*, *56*, 633–669. https://doi.org/10.1111/j.1744-6570.2003.tb00753.x

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*, 407–434. https://doi.org/10.1177/0018720814547570

Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in automation. *IEEE Intelligent Systems*, *28*, 84–88. https://doi.org/10.1109/MIS.2013.24

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Taylor & Francis.

Howell, P. D. C. (2011). *Statistical methods for psychology* (8th ed.). Wadsworth Publishing.

Iio, T., Shiomi, M., Shinozawa, K., Shimohara, K., Miki, M., & Hagita, N. (2015). Lexical entrainment in human robot interaction. *International Journal of Social Robotics*, *7*, 253–263. https://doi.org/10.1007/s12369-014-0255-x

Johnson, C. J., Demir, M., Zabala, G. M., He, H., Grimm, D. A., Radigan, C., Wolff, A. T., Cooke, N. J., McNeese, N. J., & Gorman, J. C. (2020, August). Training and verbal communications in human-autonomy teaming under degraded conditions. In *2020 IEEE conference on cognitive and computational aspects of situation management (CogSIMA)* (pp. 53–58). IEEE. https://doi.org/10.1109/CogSIMA49017.2020.9216061

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems*, *2*, 26–41. https://doi.org/10.1145/357417.357420

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*, 91–95. https://doi.org/10.1109/MIS.2004.74

Körber, M. (2019). Theoretical considerations and development of a questionnaire to measure trust in automation. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Advances in intelligent systems and computing: Vol. 823 Proceedings of the 20th Congress of the International Ergonomics Association* (pp. 13–30). Springer, Cham.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*, 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Lematta, G. J., Corral, C. C., Buchanan, V., Johnson, C. J., Mudigonda, A., Scholcover, F., Wong, M. E., Ezenyilimba, A., Baeriswyl, M., Kim, J., Holder, E., Chiou, E. K., & Cooke, N. J. (2021). Remote research methods for Human–AI–Robot teaming. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 1–18.

MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: The hidden cost of team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 61–82). American Psychological Association.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*, 277–301. https://doi.org/10.1080/14639220500337708

Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology*, *87*, 3–13. https://doi.org/10.1037/0021-9010.87.1.3

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*, 273–283. https://doi.org/10.1037/0021-9010.85.2.273

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*, 709–734. https://doi.org/10.5465/amr.1995.9508080335

Mayer, R. C., & Gavin, M. B. (2005). Trust in management and performance: Who minds the shop while the employees watch the boss? *Academy of Management Journal*, *48*, 874–888. https://doi.org/10.5465/amj.2005.18803928

Maynard, M. T., Kennedy, D. M., & Sommer, S. A. (2015). Team adaptation: A fifteen-year synthesis (1998–2013) and framework for how this literature needs to "adapt" going forward. *European Journal of Work and Organizational Psychology*, *24*, 652–677. https://doi.org/10.1080/1359432X.2014.1001376

McGrath, J. E. (1990). Time matters in groups. In J. Galegher, R. E. Kraut, & C. Egido (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work* (pp. 23–61). Lawrence Erlbaum Associates.

McNeese, N., Demir, M., Chiou, E., Cooke, N., & Yanikian, G. (2019, January). Understanding the role of trust in human-autonomy teaming. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. https://doi.org/10.24251/HICSS.2019.032

McNeese, N. J., Demir, M., Chiou, E. K., & Cooke, N. J. (2021). Trust and team performance in human–autonomy teaming. *International Journal of Electronic Commerce*, *25*, 51–72. https://doi.org/10.1080/10864415.2021.1846854

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors*, *60*, 262–273. https://doi.org/10.1177/0018720817743223

Meier, U. (2006). A note on the power of Fisher's least significant difference procedure. *Pharmaceutical Statistics*, *5*, 253–263. https://doi.org/10.1002/pst.210

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct.

Journal of Management, *36*, 876–910. https://doi.org/10.1177/0149206309356804

Myers, C., Ball, J., Cooke, N., Freiman, M., Caisse, M., Rodgers, S., Demir, M., & McNeese, N. (2018). Autonomous intelligent agents for team training. *IEEE Intelligent Systems*, *34*, 3–14. https://doi.org/10.1109/MIS.2018.2886670

O'Neill, T., McNeese, N., Barron, A., & Schelble, B. (2020). Human–autonomy teamingteamingAutonomy Teaming: A review and analysis of the empirical literature. *Human factors*, *Advance online publication*. https://doi.org/10.1518/001872008X375009

Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors*, *50*, 903–933. https://doi.org/10.1518/001872008X375009

Salas, E., & Fiore, S. M. (Eds). (2004). *Team cognition: Understanding the factors that drive process and performance*. American Psychological Association. https://doi.org/10.1037/10690-000.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small Group Research*, *36*, 555–599. https://doi.org/10.1177/1046496405277134

Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006). Does crew resource management training work? An update, an extension, and some critical needs. *Human Factors*, *48*, 392–412. https://doi.org/10.1518/001872006777724444

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, *58*, 377–400. https://doi.org/10.1177/0018720816634228

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

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT press.

Walliser, J. C., de Visser, E. J., Wiese, E., & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, *13*, 258–278. https://doi.org/10.1177/1555343419867563

Woods, D. D. (2016). The risks of autonomy: Doyle's catch. *Journal of Cognitive Engineering and Decision Making*, *10*, 131–133. https://doi.org/10.1177/1555343416653562

Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI- assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295–305. https://doi.org/10.1145/3351095.3372852

Craig J. Johnson is a human systems engineering PhD student 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 Master of Science in human systems engineering from Arizona State University.

Mustafa Demir is currently a senior quantitative researcher working at Ira. A. Fulton Schools of Engineering at Arizona State University. He received his PhD in simulation, modeling, and applied cognitive science, focusing on team coordination dynamics in human–autonomy teaming from Arizona State University in Spring 2017.

Nathan J. McNeese is an assistant professor and the director of the Team Research Analytics in Computational Environments Research Group in the School of Computing at Clemson University. He received his PhD in information sciences and technology with a focus on team decision-making, cognition, and computer-supported collaborative work from The Pennsylvania State University in 2014.

Jamie C. Gorman is an associate professor in engineering psychology at the Georgia Institute of Technology. He received his PhD from New Mexico State University in psychology in 2006.

Alexandra T. Wolff is a human systems engineering master's student and graduate research assistant at Ira A. Fulton Schools of Engineering at Arizona State University. She holds a Bachelor of Science in human systems engineering from Arizona State University.

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. She received her PhD in cognitive psychology from New Mexico State University in 1987.

Date received: January 22, 2021
Date accepted: August 30, 2021