Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming

Article *in* Human Factors The Journal of the Human Factors and Ergonomics Society · November 2017

DOI: 10.1177/0018720817743223

4 authors:

Nathan J. McNeese
Clemson University

Mustafa Demir
Arizona State University

Nancy J Cooke
Arizona State University

Christopher W. Myers
Air Force Research Laboratory

Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming

Nathan J. McNeese, Clemson University, South Carolina, **Mustafa Demir**, **Nancy J. Cooke**, Arizona State University, Mesa, and **Christopher Myers**, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio

Objective: Three different team configurations are compared with the goal of better understanding human-autonomy teaming (HAT).

Background: Although an extensive literature on human-automation interaction exists, much less is known about HAT in which humans and autonomous agents interact as coordinated units. Further research must be conducted to better understand how all-human teams compare to HAT.

Methods: In an unmanned aerial system (UAS) context, a comparison was made among three types of three-member teams: (1) synthetic teams in which the pilot role is assigned to a synthetic teammate, (2) control teams in which the pilot was an inexperienced human, and (3) experimenter teams in which an experimenter served as an experienced pilot. Ten of each type of team participated. Measures of team performance, target processing efficiency, team situation awareness, and team verbal behaviors were analyzed.

Results: Synthetic teams performed as well at the mission level as control (all human) teams but processed targets less efficiently. Experimenter teams performed better across all other measures compared to control and synthetic teams.

Conclusion: Though there is potential for a synthetic agent to function as a full-fledged teammate, further advances in autonomy are needed to improve team-level dynamics in HAT teams.

Application: This research contributes to our understanding of how to make autonomy a good team player.

Keywords: human-autonomy teaming, synthetic agent, teamwork, team cognition

Address correspondence to Nathan J. McNeese, Human-Centered Computing, Clemson University, 218 McAdams Hall, Clemson, SC 29631, USA; e-mail: mcneese@clemson.edu.

HUMAN FACTORS
Vol. XX, No. X, Month XXXX, pp. 1–12
DOI: 10.1177/0018720817743223

INTRODUCTION

What Does Human-Autonomy Teaming Mean?

Teams have long been used to complete work in a variety of tasks and contexts (Salas, Cooke, & Rosen, 2008). Traditionally, teamwork has been defined as two or more humans interdependently working toward a common goal (Salas, Dickinson, Converse, & Tannenbaum, 1992). Most of our knowledge pertaining to teamwork is founded on the basis of human-human interactions, including our understanding of situation awareness (Gorman, Cooke, & Winner, 2006), teamwork and taskwork knowledge (Lim & Klein, 2006), transactive memory (Moreland, 1999), and team cognition (Cooke, 2015; Cooke, Gorman, Myers, & Duran, 2013) in all-human teams. Yet, in recent years, autonomy has become intelligent enough to be considered a teammate, as opposed to a servant. We refer to this concept as *human-autonomy teaming* (HAT).

Note that we refer to this as human-*autonomy* teaming and not human-*automation* teaming. Autonomy capitalizes on technology's ability to make intelligent decisions and adapt to task, situation, and context, thus allowing it to improve on its own performance over time (Cox, 2013). Autonomy encompasses

> systems which have a set of intelligence-based capabilities that allow it to respond to situations that were not programmed or anticipated in the design (i.e., decision-based responses). Autonomous systems have a degree of self-government and self-directed behavior (with the human's proxy for decisions). (USAF, 2013, p. 3)

For the purpose of this article, we define automation as technology that requires human intervention or control and autonomy as technology capable of working alongside humans as teammates, carrying out the essential taskwork and teamwork functions of a human teammate. In the research presented here, we identify the essential teamwork functions necessary for effective HAT.

Much of the work prior to HAT comes from the area of human-automation interaction. Both human-robot teaming (Gao, Cummings, & Solovey, 2016; Shah & Breazeal, 2010) and human supervisory control of automation (Sheridan, 2012) have investigated how humans interact with lower levels of autonomy. Chen and Barnes (2014) present a comprehensive review of issues pertaining to human-automation interaction and human-agent teaming in the context of robotics. Recent research has focused on developing robots that effectively work with humans (Zhang, Narayanan, Chakraborti, & Kambhampati, 2015), the interactions among humans and robots during teaming (Bartlett & Cooke, 2015), and modeling human cognition in robots (Goodrich & Yi, 2013).

Essential Teamwork Functions for Autonomous Teammates

High levels of autonomy are necessary to take on the role of a human teammate in complex situations. First, one must understand the task at hand and take on a distinct role on the team. Second, there is a need for the teammate to interact with other teammates—to communicate. These would appear to be the essential requirements of an autonomous teammate. But are they sufficient? From the literature, we know that effective human teammates exhibit other team behaviors such as backup behaviors (knowing other team member responsibilities) (Salas, Sims, & Burke, 2005) and providing feedback (response based on team behavior) (McIntyre & Salas, 1995). Additional research has indicated that effective human teamwork is accompanied by good coordination or getting the right information to the appropriate teammate at the correct time (Cooke et al., 2013). Effective human teammates anticipate information needs of teammates and "push" that information out at a time when it is needed. This information is contrasted to the "pulling" of information from teammates (i.e., asking for the information). Can HAT function adequately when the autonomous teammate lacks some of these subtle coordination behaviors exhibited by humans?

The Current Study

In this article, we present results of an experiment in which an autonomous synthetic teammate interacted with two human teammates in the context of an unmanned aerial system (UAS) task with the goal of photographing ground targets.

The synthetic agent was capable of performing the task of the pilot and communicated with human teammates using text chat but lacked coordination skills. Teams working with a synthetic teammate were compared to teams of all humans and teams with an experienced human pilot. Specifically, a comparison was made among three types of three-member teams: (1) synthetic teams in which the pilot role is assigned to a synthetic teammate, (2) control teams in which the pilot was an inexperienced human, and (3) experimenter teams in which an experimenter served as an experienced pilot. The experimenter condition was conceived as a method to establish a baseline for teams with a synthetic agent operating at expert levels. Ten of each type of team participated. Measures of team performance, target processing efficiency, team situation awareness, and team verbal behaviors were analyzed.

We hypothesized that teams with a synthetic teammate would demonstrate poorer team coordination and thus poorer target processing and overall performance compared to the control and experimenter conditions. We further hypothesized that experimenter teams would display better coordination and thus better target processing and performance than control teams due to coordination prompts of the experimenter. In general, our study compares HAT to all-human teams to: (a) evaluate the impact of autonomy on team effectiveness and (b) through revealing the shortcomings of HAT, understand what teamwork skills are essential for autonomy.

METHOD

Participants

Seventy participants (both graduate and undergraduate) were recruited from a large Southwestern university, resulting in 30 teams who completed the experiment. Three participants per team in the control condition and two participants per team (one each for the photographer and navigator roles) in both the experimenter and synthetic conditions were recruited. The pilot role was filled by a trained confederate in the experimenter condition (i.e., the same confederate was present in all experimental trials) or a synthetic teammate in the synthetic condition. In the synthetic condition, participants were aware that the pilot was a synthetic agent. Participants were required to be fluent in English and have normal or corrected-to-normal vision. Ages ranged from 18 to 38 years (M_{age} = 23.7, SD_{age} = 3.3) across 60 males and 10 females. Every team completed one eight-hour session, for which each individual was compensated $10 per hour.

Task and Roles

The Cognitive Engineering Research on Team Tasks Unmanned Aerial System–Synthetic Task Environment (CERTT UAS-STE) provided the context for this study. The UAS-STE is based on the United States Air Force Predator UAS ground control station. The UAS-STE task requires three different, interdependent teammates, each with a unique role relevant to the team's objective of efficiently taking good photos of target waypoints: (1) pilot—controls the UAS's heading, altitude, and airspeed following the flight plan; (2) navigator—provides a dynamic flight plan as well as speed and altitude restrictions; and (3) photographer—monitors sensor equipment and takes photographs in accord with UAS airspeed and altitude (Cooke & Shope, 2004). Based on task and team roles, an ideal team interaction at a target (to be photographed) waypoint consists of three main communication events: (1) The navigator sends information about the target to the pilot, (2) the pilot negotiates with the photographer about the target's altitude and airspeed restrictions with the goal of correctly adjusting camera settings, and (3) the photographer provides feedback on whether or not a good photo was acquired (Cooke et al., 2007). This interaction is critical and depends on all roles.

Spatial characteristics of the task scenarios include terrain features, weather obstructions, enemy activity, target priority, and restricted operating zones (ROZ) location. These characteristics are modified via a waypoint library that defines the scenario in terms of a world of waypoints, each with varying features (Cooke, Rivera, Shope, & Caukwell, 1999). Target waypoints are waypoints at which the photographer is able to take photographs and are located within the ROZ with prescribed entry and exit points. In addition, target waypoints consist of restrictions in airspeed, altitude, and effective radius, resulting in the need for communication and coordination of information throughout the team.

In the current study, the goals of the five 40-minute missions were the same (i.e., identifying and processing the targets). However, missions differed by numbers and locations of target waypoints with Missions 1 through 4 being of comparable complexity with 11 to 13 targets in each and Mission 5 being a more difficult mission with 20 targets. Learning could be demonstrated across Missions 1 through 4, with Mission 5 being a stringent test of team performance (Cooke et al., 2007).

Throughout the missions, participants would be alerted to new targets (targets not on the navigator's list). These unexpected targets or "roadblocks" provided an opportunity for measuring the team's situation awareness. Unexpected target waypoints were provided by the experimenters to the navigator when the team entered a specific simulated zone in each mission. For example, when a team gets close to the unexpected target waypoint of "WP8," the experimenter sends the following information to the navigator: "Enemy forces have just left target WP8. You are clear to enter area and take a photo." There were three roadblocks in Missions 1 and 2, four in Missions 3 and 4, and five in Mission 5.

During each mission, teams were told to obtain as many "good" photos of targets as possible while avoiding alarms and rule violations. A "good" photograph is one that corresponded with the photographer's library of good photos. Only using the proper camera settings would result in a good photo. Missions terminated either after 40 minutes had elapsed or when team members believed that the mission goals—taking a good photo for each target—had been completed.

TABLE 1: Example Team Interaction in the Unmanned Aerial Vehicle Synthetic Task Environment

Event	Sender	Receiver	Utterance
Information	Navigator	Pilot	The next waypoint is H-Area. It is a target. The airspeed restriction is 50 to 200. There is no altitude restriction. The effective radius is 5.
Negotiation	Pilot	Photographer	The target altitude for H-Area is 2000.
			The target airspeed for H-Area is 190.
Feedback	Photographer	Pilot and navigator	Got the photo. Let's go.

The Synthetic Teammate

As part of a separate project, a synthetic teammate (Ball et al., 2010) was developed over the past decade that is capable of serving as a full-fledged synthetic teammate in the pilot role of a three-agent UAS ground control crew in the CERTT UAS-STE. The synthetic teammate was developed using the ACT-R cognitive modeling architecture (Anderson, 2007) and interacts with human teammates using text chat. The synthetic teammate is "full-fledged" in the sense that it does not perform only part of a teammate's task (e.g., a decision aid, or route planner, or warning checker, etc.). It is responsible for all aspects of the task and must integrate with the other team members just as they must integrate with the synthetic teammate. That is to say, the synthetic teammate cannot be set aside and the team expect to perform the task well. The synthetic teammate is a critical component on the team, just like its human counterparts.

Although the synthetic teammate is autonomous in that it can choose its own course of actions based on its experiences during the task and a dynamic situation representation, it was not developed with explicit teamwork skills. For example, the synthetic teammate does not have a concept of coordination—the timely sharing of task-critical information. However, it does contain a dialog management system that understands requests for information and when it needs to ask for information. This experiment tests the importance of these additional teamwork skills to the extent that the synthetic teammate succeeds or fails.

During training, participants in the synthetic condition were informed that the pilot was a synthetic agent and were instructed to communicate with the synthetic teammate using clear text messages (i.e., not misspelled or cryptic).

An example of a good communication sequence between the synthetic teammate and human team members is provided in Table 1. In this specific example, the synthetic teammate (a pilot) asks the navigator for information regarding an upcoming target waypoint called H-Area. Once the navigator supplies clear information, the synthetic teammate then provides airspeed and altitude information to the photographer. Finally, if the waypoint was the target, the synthetic teammate would wait for feedback from the photographer as to whether the team acquired a good photo.

Materials and Equipment

Within the CERTT UAS-STE, human participants use a role-specific console with two embedded computers, allowing the participant to monitor the alarms and warnings for role-specific work and send messages via the chat window to the other two roles (Figure 1). Role-specific teamwork was conducted through one of the console's computers (including two touchscreens, one keyboard, and one mouse). The console's other computer (including one touchscreen and one keyboard) was used for text communication with the other two team members. Throughout the experiment, the navigator and photographer were seated in a room and separated by partitions that prevented face-to-face contact. The pilot was alone in another room.

An additional computer was also used to operate the synthetic teammate. This computer ran the synthetic teammate's software and communicated with the participant consoles using text chat.

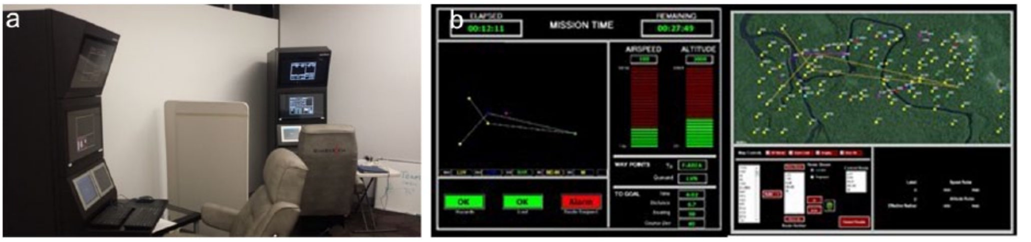

Figure 1. The Cognitive Engineering Research on Team Tasks Unmanned Aerial System–Synthetic Task Environment consoles and a view from the navigator screen. (a) Participant consoles. (b) Example screen view from the navigator console.

Procedure

During the experimental session, participants first signed an informed consent document and were then randomly assigned to their roles (pilot [depending on condition], navigator, photographer).

Before the first mission, participants received 30 minutes of role-specific PowerPoint training. After the PowerPoint training was completed, participants (in all conditions) were then given 30 minutes of hands-on practice conducting their specific tasks as a team. In the synthetic condition, the participants were specifically trained on how to communicate with the synthetic teammate. Because interacting with the synthetic teammate is different than interacting with a human, more specialized communication training was provided to these teams. Teams in other conditions did not receive specific communication training other than what was included in their normal training. In the experimenter condition, the pilot used a coordination script that specified conditions under which specific information needed to be requested because it was not given in a timely manner (e.g., if the navigator failed to communicate target restrictions within one minute, then the pilot requested this information). After the hands-on training ended, participants were given a 15-minute rest break.

Once training ended, the experiment began with each team photographing targets during each of the five 40-minute missions for the duration of the eight-hour experimental session (see Table 2). Participants were given a 15-minute rest break after Mission 1, a 30-minute lunch break after Mission 2, and then 15-minute rest breaks after Missions 3, 4, and 5.

TABLE 2: Experimental Session

Sessions
1. Consent forms
2. PowerPoint training
3. Hands-on training
4. Mission 1
5. NASA TLX/knowledge
6. Mission 2
7. Mission 3
8. Mission 4
9. Mission 5
10. NASA TLX/knowledge
11. Demographic questions/debriefing
12. Post checklist by the experimenters

In two separate sessions, after Missions 1 and 5, taskwork knowledge, teamwork knowledge, and workload (NASA TLX; Hart & Staveland, 1988) were measured to examine possible changes in knowledge and workload over the experimental session. Lastly, demographic information was collected, and participants were debriefed on the purpose of the experiment. Knowledge measures, NASA TLX, and demographic measures were all analyzed but did not result in any significant differences among conditions or across missions and are not reported here.

Measures

Several measures were taken in this experiment, including: mission-level team performance, target processing efficiency, and team process measures, including process ratings, verbal behaviors, and team situation awareness.

TABLE 3: Team Verbal Behaviors

Behaviors	Push/Pull	Description
General status updates	Push	Informing other team members about current status
Repeated requests	Pull	Requesting the same information or action from other team member(s)
Inquiry about status of others	Pull	Inquiring about current status of others and expressing concerns
Suggestions	Push	Making suggestions to the other team members
Planning ahead	Push	Anticipating next steps and creating rules for future encounters
Positive communication	NA	Helping out team members by providing information and acknowledgement of member's speech
Negative communication	NA	Argument among the team members due to conflicting goals or incorrect destination
Unclear communications	NA	Sending information with misspellings and ambiguous terms that experimenters cannot understand

Team performance. Team performance, an outcome-based measure of team effectiveness, is the weighted composite of team-level mission parameters, including time spent in warning and alarm states, number of missed targets, and rate of good target photographs per minute (which was weighted most heavily among the parameters). Teams begin each mission with a score of 1,000, and points are deducted based on the final values of the mission subscores (see Supplementary Material on the *HF* Web site for weighted calculations of the team performance score and its subscores; Cooke et al., 2007).

Target processing efficiency. Target processing efficiency takes into account the time spent inside a target waypoint to get a good photo (higher scores equate to more efficiency). Teams can get 1,000 points possible at each target level. Each score is deducted with the number of seconds spent in the target radius and 200 penalty points for missed photos (Cooke et al., 2007). In comparison to the team performance score, the target processing efficiency score provides a specific score oriented only to targets.

Team process: Process ratings. During the experiment, two experimenters rated the quality of team process behaviors for each target after it was photographed by the teams. Experimenters rated process at each target in each mission independently based on three dimensions (ratings ranging from 1 = *poor* and 5 = *excellent*): (1) *coordination*: communicating with the correct team member about the correct information in the correct sequence; (2) *timeliness of the coordination*: based on the team's ability to coordinate through relevant interaction in time to process the target (requires looking at when the interactions occurred and the UAS's relative position to the target); and (3) *quality of communication*: based on how clear and unique the communications are (ideally, minimizing the need for repetition).

Team process: Verbal behaviors. Eight verbal behaviors were identified from previous CERTT UAS-STE data as being associated with team effectiveness (Table 3). Instances of these behaviors were tagged by two experimenters and counted for each mission. Five of these behaviors were classified as pushing information and pulling information among the team members. In this task, more pushing or anticipating information needs is associated with good teamwork (Demir, McNeese, & Cooke, 2017).

Team process: Situation awareness. Teams had to respond to occasional "roadblocks," such as the introduction of a new target waypoint. The degree to which the team members took action and overcame the roadblocks was a measure of team situation awareness. To overcome a roadblock, the new target needs to be successfully

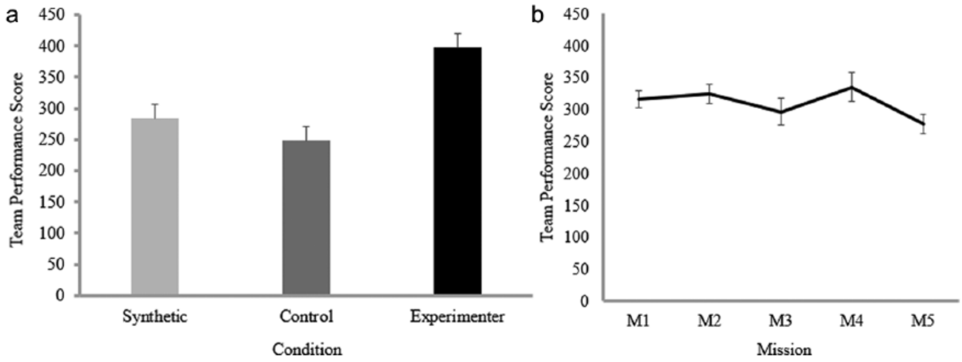

Figure 2. Team performance (a) across the conditions (Synthetic = Control < Experimenter) and (b) across the missions. Error bars provide the standard error of the mean.

photographed (Gorman, Cooke, & Amazeen, 2010). In this study, we counted the number of completed roadblocks in each mission.

RESULTS

Team Performance

We performed a two-factor split-plot analysis of variance (ANOVA) to determine whether the conditions differed with respect to their team performance scores over missions. Mauchly's test indicates that the assumption of sphericity for the repeated measure (mission) was not satisfied, $\chi^2(9) = 22.1$, $p < .05$, $\varepsilon = .76$. Therefore, degrees of freedom were corrected using the Greenhouse-Geisser correction. In this case, SPSS uses the epsilon value to implement a larger adjustment on the probability value (and thus, a more conservative approach to deal with false positives). The ANOVA results indicate that the condition main effect, $F(2, 27) = 11.5$, $MS_e = 26{,}337$, $p < .001$, $\eta^2 = .46$, and the mission main effect, $F(3.05, 85.3) = 2.99$, $MS_e = 6956$, $p < .05$, $\eta^2 = .10$, were statistically significant. However, the condition by mission interaction effect was not significant, $F(8, 108) = 1.88$, $MS_e = 5{,}297$, $p = .07$.

The pairwise test results indicate that the experimenter teams ($M_{Exp} = 398$, $SD_{Exp} = 94.9$, $p < .05$) performed better than the synthetic and control teams, ($M_{Syn} = 295$, $SD_{Syn} = 105$, $M_{Cont} = 249$, $SD_{Cont} = 95.3$, $p < .05$). The synthetic teams' performance was not significantly different from the control teams' performance ($p = .16$; Figure 2).

The significant mission main effect indicates that overall team performance across four missions ($M_{Mission1} = 316$, $SD_{Mission1} = 85.7$, $M_{Mission2} = 325$, $SD_{Mission2} = 97.8$, $M_{Mission3} = 297$, $SD_{Mission3} = 140$, $M_{Mission4} = 336$, $SD_{Mission4} = 145$) was not significantly different but that team performance was significantly reduced during the last mission ($M_{Mission5} = 278$, $SD_{Mission5} = 101$) compared to Missions 1, 2, and 4, $p < .05$. This pattern is not unexpected given that the last mission included more targets and thus was more difficult than the first four.

Target Processing Efficiency

The teams' target processing efficiency was analyzed via a repeated measure three-factor mixed ANOVA with condition as a between-teams manipulation and mission and target nested within missions as within teams factors. Table 4 summarizes the results of the mixed ANOVA.

There was a significant main effect of condition, $F(2, 32.2) = 10.9$, $MSe = 47{,}398$, $p < .001$. Pairwise tests indicate that the synthetic teams had significantly poorer target processing efficiency than the control and experimenter teams, $M_{Syn} = 820$, $SD_{Syn} = 158$, $M_{Cont} = 866$, $SD_{Cont} = 120$, $M_{Exp} = 939$, $SD_{Exp} = 58.1$, $p < .001$, and the experimenter teams were more efficient than control and synthetic teams ($p < .05$; Figure 3).

The significant mission effect indicates an improvement in overall efficiency from Mission 3 to Mission 4 ($M_{Mission4} = 882$, $SD_{Mission4} = 109$, $p = .07$) and from Mission 4 to Mission 5

TABLE 4: The ANOVA Results for Target Processing Efficiency

Source	*df*	*F*	*p*	η^2
Condition	2	10.85	.00	.40
Condition by target (within mission)	80	1.41	.01	.13
Mission	4	2.13	.08	.03
Condition by mission	8	1.27	.26	.05
Target (within mission)	54	2.40	.00	.14

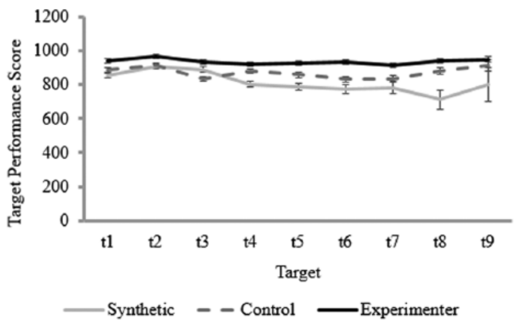

Figure 3. Target processing efficiency across the conditions (Synthetic < Control < Experimenter) and across the targets. Error bars provide the standard error of the mean.

($M_{Mission5}$ = 917, $SD_{Mission5}$ = 94.5, $p < .05$). The significant target (within mission) effect indicates that target processing efficiency increased during the first two targets, and then it was significantly decreased until Target 6. After that, there was an increment again on target processing efficiency until the last target.

These main effects should be interpreted in the context of the significant condition by target (within mission) interaction that indicates that the experimenter and control teams' target processing efficiency was consistent across targets, whereas the synthetic teams' target processing efficiency decreased over time and targets (see Figure 3).

Performance Score: Summary

Synthetic teams performed as well as control teams on the outcome-based measure of team performance, though the experimenter outperformed both. However, the synthetic teams did not process targets as efficiently as control teams, who were less efficient than experimenter teams. Target processing efficiency emphasizes team coordination. We turn next to our process measures to better understand the synthetic team deficits.

Team Process: Process Ratings

A weighted Cohen's κ was conducted to determine if there was an agreement between two experimenters' observations on recording the team process ratings. There was only fair agreement between the two experimenters' observation, κ = .35 (95% CI [.29-.40]), $p < .001$, and therefore, we do not report the detailed results.

Team Process: Team Verbal Behaviors

Cohen's κ was conducted to determine if there was an agreement between two experimenters' observations on recording the teams' verbal behaviors, and it was found that there was substantial agreement between the two experimenters' observations, κ = .77 (95% CI [.75-.79]), $p < .001$. In this case, we took the mean of both experimenters' observations of the verbal behavior and then proceeded with the following analysis on verbal behavior.

To analyze the number of each of the verbal behaviors, we performed a 3 (condition) × 5 (mission) × 3 (role) × 8 (verbal behavior) repeated measures Multivariate ANOVA on the counts of eight team verbal behaviors for each role. Mauchly's test indicates that the assumption of sphericity for the repeated measure (mission) was satisfied, $\chi^2(9) = 12.9$, $p = .17$. However, the assumption of sphericity was not satisfied for role, $\chi^2(2) = 10.7$, $p < .001$, $\varepsilon = .75$; behavior, $\chi^2(27) = 124$, $p < .001$, $\varepsilon = .58$; role by mission interaction, $\chi^2(35) = 96.8$, $p < .001$, $\varepsilon = .48$; or role by behavior interaction, $\chi^2(104) = 439$, $p < .001$, $\varepsilon = .30$. Therefore, degrees of freedom were corrected using the Greenhouse-Geisser correction and reported in Table 5. Table 5 presents the results of the MANOVA. There are significant three-way and four-way interaction

TABLE 5: MANOVA Results for Team Verbal Behaviors

Source	*df*	*F*	*p*	η^2
Mission	4	4.49	.002	.14
Mission by condition	8	1.30	.250	.09
Role	1.36	97.65	.000	.78
Role by condition	4	32.20	.000	.70
Behavior	3.42	101.51	.000	.79
Behavior by condition	14	15.35	.000	.53
Mission by role	3.83	2.15	.083	.07
Mission by role by condition	16	1.97	.016	.13
Mission by behavior	28	3.31	.000	.11
Mission by behavior by condition	56	1.76	.001	.12
Role by behavior	4.33	30.52	.000	.53
Role by behavior by condition	28	13.62	.000	.50
Mission by role by behavior	56	1.73	.001	.06
Mission by role by behavior by condition	112	1.38	.006	.09

effects, indicating that the number of verbal behaviors depends on the type of verbal behavior, role of the participant, mission, and condition. Due to space limitations, we summarize only the statistically significant differences between the synthetic and other conditions in Table 6.

Synthetic teams gave fewer general status updates and had more repeated requests compared to other teams. This indicates and paints a general picture of teams that are doing more pulling than pushing of information. Also, synthetic teams did not decrease inquiries to others or this pulling behavior over time, as higher performing teams did, suggesting that more implicit coordination patterns were not developing.

Team Process: Situation Awareness

We performed a two-factor split-plot ANOVA to determine whether the conditions differed with respect to the number of roadblocks overcome across missions. Mauchly's test indicates that the assumption of sphericity for the repeated measure (mission) was not satisfied, $\chi^2(9) = 18.9$, $p < .05$, $\varepsilon = .79$. Therefore, degrees of freedom were corrected using the Greenhouse-Geisser correction. The results indicate that all three effects were statistically significant: the condition by mission interaction effect, $F(8, 108) = 2.13$, $MS_e = .45$, $p < .05$, $\eta^2 = .14$; the condition effect, $F(2, 27) = 19.50$, $MS_e = 1.47$, $p < .001$, $\eta^2 = .59$; and the mission effect, $F(3.15, 85) = 11.6$, $MS_e = .45$, $p < .001$, $\eta^2 = .30$.

The experimenter teams ($M_{Exp} = 1.86$, $SD_{Exp} = .90$) overcame more roadblocks than the synthetic ($M_{Syn} = .38$, $SD_{Syn} = .60$, $p < .001$) and control teams ($M_{Cont} = .84$, $SD_{Cont} = .93$, $p < .001$). The control and synthetic teams were not statistically different ($p = .069$). Also, the number of roadblocks overcome generally increased across the first three missions: Missions 1 to 2 ($M_{Mission1} = .50$, $SD_{Mission1} = .73$, $M_{Mission2} = .80$, $SD_{Mission2} = .85$, $p < .05$) and Missions 2 to 3 ($M_{Mission3} = 1.37$, $SD_{Mission3} = 1.13$, $p < .001$) (Figure 4).

These main effects should be interpreted in the context of the significant condition by mission interaction, which indicates that the improvement in number of roadblocks overcome in the early missions occurs only for the control and experimenter teams but not the synthetic teams.

Overview of Results

Table 7 summarizes the results across conditions. In general, synthetic teams performed comparable to control teams. However, the synthetic teams processed targets less efficiently and tended to pull information more than push it.

TABLE 6: Pairwise Comparison Results: Verbal Behavior Differences Between Synthetic and Other Teams

Behaviors	Results
General status updates	Between the team members were more frequent in the experimenter condition over time, whereas the general status updates were less frequent in synthetic (M_{Exp} = 8.84, M_{Syn} = 4.87, p < .001) and control conditions (M_{Exp} = 8.84, M_{Cont} = 6.79, p < .05).
Inquiries to others	Decreased across the missions for the control (M_1 = 4.88, M_5 = 2.93, p < .05) and the experimenter conditions (M_1 = 6.08, M_5 = 4.25, p < .05). The synthetic teams' inquiries to others did not significantly change across the missions (M_1 = 2.07, M_5 = 2.13, p = .93).
Repeated requests	Happened more in the synthetic and control conditions than the experimenter condition (M_{Syn} = .39, M_{Cont} = .25, M_{Exp} = .22).

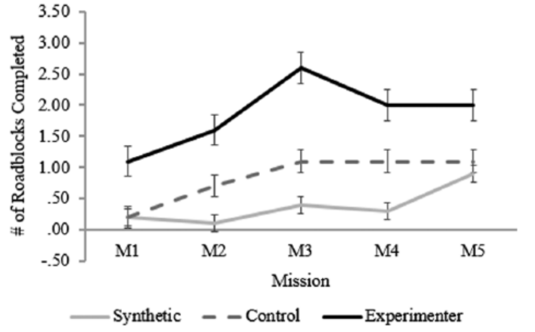

Figure 4. Estimated means of proportion of overcoming the roadblocks after triggered (Synthetic = Control < Experimenter). Error bars provide the standard error of the mean.

DISCUSSION AND CONCLUSION

The evaluation of an autonomous synthetic agent that acts as a full-fledged teammate in a team context represents an advance in human-autonomy teaming. The synthetic teams in this study show both promise and limitations. When one compares team performance across conditions, synthetic teams perform as well at the mission level as all-human teams. This bodes well for the future of HAT in contexts such as the UAS one used here, but there are caveats. Even though synthetic teams performed well, there were limited roadblocks per mission, and they were not of high levels of difficulty. That is, the roadblocks were simply the addition of new ad hoc targets rather than failures in system displays, controls, or communication breakdowns. Research is ongoing to address the response of synthetic teams to failures and the resilience in their aftermath.

At the same time, the synthetic teams processed targets less efficiently than their all-human counterparts. Target processing efficiency relies heavily on the timely coordination of the three teammates at target waypoints. Certain information has to be passed by certain people to certain others at certain times. Good teams become adept at this repeated coordination process, often "front-loading" information in anticipation of upcoming targets. Examination of the verbal behaviors suggested that there was less anticipation of information needs (i.e., pushing) for synthetic teams as opposed to other teams. The synthetic teammate was capable of communications and understood its information needs but did not seem to understand the information needs of its fellow teammates. Interestingly, this lack of anticipation seemed to be contagious as the entire team did less pushing. Optimal team interaction patterns should consist of less pulling and more pushing of information as teams develop (i.e., the opposite found in the synthetic teams). This seems to indicate that synthetic teams had difficulty learning how to develop an effective coordination strategy. This further indicates a weakness of the synthetic teammate that should be relatively easy to overcome in future iterations. This also highlights the importance of some subtle teamwork behaviors that seem to come naturally for humans.

Experimenter teams achieved better team performance across the board. This is a promising

TABLE 7: Summary of Condition Effects

Measures	Result Highlights
Team performance	Synthetic = Control < Experimenter
Target processing efficiency	Synthetic < Control < Experimenter
Team verbal behavior	Synthetic teams pull more than push
Team situation awareness	Synthetic = Control < Experimenter

finding, as it demonstrates (1) the future potential of a synthetic teammate further developed to operate with the capabilities of an "experienced" pilot within the UAV-STE and (2) what can currently be achieved by inserting an "experienced" synthetic teammate into a team training exercise. In this scenario, training occurs implicitly through the deliberate and timely pushing and pulling of information by the "experienced" agent. This type of training has the ability to accelerate training team coordination through synthetic teammates that push and pull information in an efficient manner.

In conclusion, we have demonstrated that advancements of both autonomous technology and human's ability to interact with it in a meaningful way are critical to the success of teamwork in a UAS context. The human-autonomy teams in this study show promise. This work suggests that depending on the task and context, synthetic teammates may serve as replacements of human teammates for team training delivery. As human-autonomy teams become more prevalent and are realized in different contexts, it is important that the research community continues to develop effective synthetic teammates and seeks to understand how humans best interact with them. There is much work that needs to be conducted, specifically better understanding the limitations of HAT and how to overcome those limitations, especially in more complex and dynamic environments than the one used here. The work presented here is one of the first steps in understanding real HAT.

ACKNOWLEDGMENTS

This research was partially supported by ONR Award N000141110844 (Program Managers: Marc Steinberg, Paul Bello). We also acknowledge the assistance of Jerry Ball, Mary Friedman, Steven M. Shope, Cade Bartlett, Ashley Knobloch, and Lourdes Reyes.

KEY POINTS

- As machines grow in levels of intelligence, humans can transition from supervisory control of machines to human-autonomy teaming.
- In this study, teams of humans and a synthetic teammate performed comparably to all-human teams at some levels, though processed targets less efficiently, anticipating information needs less than teams of all humans.
- Though there are some limitations of synthetic teammates, the limits do not seem insurmountable such that in the future they may serve as replacements of human teammates for team training delivery and eventually real-world missions.

SUPPLEMENTARY MATERIAL

The online supplementary material is available with the manuscript on the *HF* Web site.

REFERENCES

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford, UK: Oxford University Press.

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*(3), 271–299.

Bartlett, C. E., & Cooke, N. J. (2015). Human-robot teaming in urban search and rescue. *Proceedings of the Human Factors and Ergonomics Society 59th Annual Meeting* (pp. 250–254). Santa Monica, CA: Human Factors and Ergonomics Society.

Chen, J. Y., & Barnes, M. J. (2014). Human-agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, *44*(1), 13–29.

Cooke, N. J. (2015). Team cognition as interaction. *Current Directions in Psychological Science*, *24*(6), 415–419.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–285.

Cooke, N. J., Gorman, J., Pedersen, H., Winner, J., Duran, J., Taylor, A., . . . Rowe, L. (2007). *Acquisition and retention of team coordination in command-and-control*. Retrieved from http://www.dtic.mil/docs/citations/ADA475567

Cooke, N. J., Rivera, K., Shope, S. M., & Caukwell, S. (1999). A synthetic task environment for team cognition research. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 303–308). Santa Monica, CA: Human Factors and Ergonomics Society.

Cooke, N. J., & Shope, S. M. (2004). Synthetic task environments for teams: CERTT's UAV-STE. In N. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. Hendrick (Eds.), *Handbook of human factors and ergonomics methods* (pp. 41–46). Boca Raton, FL: CRC Press.

Cox, M. T. (2013). *Goal-driven autonomy and question-based problem recognition*. Retrieved from http://www.cogsys.org/papers/2013poster9.pdf

Demir, M., McNeese, N. J., & Cooke, N. J. (2017). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, *46*, 3–12.

Gao, F., Cummings, M. L., & Solovey, E. (2016). Designing for robust and effective teamwork in human-agent teams. In R. Mittu, D. Sofge, A. Wagner, & W. F. Lawless (Eds.), *Robust intelligence and trust in autonomous systems* (pp. 167–190). New York, NY: Springer.

Goodrich, M. A., & Yi, D. (2013). Toward task-based mental models of human-robot teaming: A Bayesian approach. In R. Shumaker (Ed.), *Virtual augmented and mixed reality. Designing and developing augmented and virtual environments* (pp. 267–276). Berlin: Springer.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010). Training adaptive teams. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*, 295–307.

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

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Lim, B.-C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*(4), 403–418.

Mcintyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. Guzzo, & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 149–203). San Francisco: Jossey-Bass.

Moreland, R. L. (1999). Transactive memory: Learning who knows what in work groups and organizations. In L. L. Thompson, J. M. Levine, & D. M. Messick (Eds.), *Shared cognition in organizations* (pp. 3–31). Hillsdale, NJ: Lawrence Erlbaum Associates.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 540–547.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex Publishing.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research*, *36*(5), 555–599.

Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human-robot teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*, 234–245.

Sheridan, T. B. (2012). Human supervisory control. G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (4th ed., pp. 990–1015). Hoboken, NJ: John Wiley.

USAF. (2013). *Autonomy science and technology strategy*. Dayton, OH: Airforce Research Laboratory.

Zhang, Y., Narayanan, V., Chakraborti, T., & Kambhampati, S. (2015). A human factors analysis of proactive support in human-robot teaming. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3586–3593). Hamburg: IEEE.

Nathan J. McNeese is an assistant professor of human-centered computing in the School of Computing at Clemson University. He received his PhD in information sciences and technology at The Pennsylvania State University in 2014.

Mustafa Demir is a postdoctoral scholar in human systems engineering at Arizona State University. He received his PhD in simulation, modeling and cognitive science at Arizona State University in 2017.

Nancy J. Cooke is a professor in human systems engineering at Arizona State University. She received her PhD in cognitive psychology from New Mexico State University in 1987.

Christopher Myers is a senior cognitive scientist at the Air Force Research Laboratory. He received his PhD in cognitive science from Rensselaer Polytechnic Institute in 2007.

Date received: August 16, 2016
Date accepted: October 16, 2017