COMMUNICATION PATTERN ANALYSIS IN HUMAN-AUTONOMY TEAMING

A Thesis
Presented to
The Academic Faculty

by

Shiwen Zhou

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
August 2022

COMMUNICATION PATTERN ANALYSIS IN HUMAN-AUTONOMY TEAMING

Approved by:

Dr. Jamie Gorman, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Nancy Cooke
Ira A. Fulton Schools of Engineering
Arizona State University

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Dr. James Roberts
School of Psychology
Georgia Institute of Technology

Date Approved: [July 18, 2022]

ACKNOWLEDGEMENTS

I would like to thank my advisor, Jamie Gorman, and the other graduate students in the lab, David Grimm, Matthew Scalia, Julie Harrison, and Terri Dunbar, for their guidance and assistance on this project. I would especially like to thank Nancy Cooke for providing the data her research team collected at Arizona State University. I would not have been able to complete this thesis without their help on the data collection portions of this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ACT-R	Adaptive Control of Thought-Rational
AI	Autonomous
AVO	Pilot
AVO0PLO	AVO sending message to PLO
AVO0DEMPC	AVO sending message to DEMPC
AVO1AVO	Communication sequence AVO, AVO
AVO1PLO	Communication sequence AVO, PLO
AVO1DEMPC	Communication sequence AVO, DEMPC
AVO2AVO	Communication sequence AVO, ( ), AVO
AVO2PLO	Communication sequence AVO, ( ), PLO
AVO2DEMPC	Communication sequence AVO, ( ), DEMPC
CERTT-RPAS-STE	Cognitive Engineering Research on Team Tasks RPAS Synthetic Task Environment
d	Cohen'd
DEMPC	Navigator
DEMPC0AVO	DEMPC sending message to AVO
DEMPC0PLO	DEMPC sending message to PLO
DEMPC1AVO	Communication sequence DEMPC, AVO
DEMPC1DEMPC	Communication sequence DEMPC, DEMPC
DEMPC1PLO	Communication sequence DEMPC, PLO
DEMPC2AVO	Communication sequence DEMPC, ( ), AVO
DEMPC2DEMPC	Communication sequence DEMPC, ( ), DEMPC
DEMPC2PLO	Communication sequence DEMPC, ( ), PLO

df	Degree of freedom
η^2_p	Partial Eta-squared
F	F-test statistic
LSA	Latent Semantic Analysis
M	Mean
p	Probability value
PLO	Photographer
PLO0AVO	PLO sending message to AVO
PLO0DEMPC	PLO sending message to DEMPC
PLO1AVO	Communication sequence PLO, AVO
PLO1DEMPC	Communication sequence PLO, DEMPC
PLO1PLO	Communication sequence PLO, PLO
PLO2AVO	Communication sequence PLO, ( ), AVO
PLO2DEMPC	Communication sequence PLO, ( ), DEMPC
PLO2PLO	Communication sequence PLO, ( ), PLO
RPA	Remotely piloted aircraft
RPAS	Remotely piloted aircraft system
SD	Standard deviation
t	Student’s t-test
TPE	Target Processing Efficiency
UAV	Unmanned aerial vehicle
%DET	Percent determinism

SUMMARY

Communication is critical to team coordination and interaction because it provides information flows allowing a team to build team cognition, which contributes to overall team performance. In recent years, autonomous (AI) team members are beginning to be considered as effective substitutes for human teammates. However, research has shown that AI team members may lack the communication skills that are required for effective team performance (McNeese et al., 2018). To better understand which aspects of communication an AI team member performs differently compared to a human team member, and how they impact team performance, the current study analyzes communication features of three-person teams that include all human teams and human-AI teams operating in a remotely piloted aircraft system (RPAS). The current study analyzed communication pattern predictability (communication determinism) and transition probabilities to measure communication flow and Latent Semantic Analysis (LSA) to measure communication content. The current study found that both communication flow and content distinguished communication in all-human teams from communication in human-AI teams and found that these communication flow and content features predicted team performance in all-human versus human-AI teams. In this way, the current study hopes these communication differences can provide feedback and suggestions to future adoption of AI as a teammate in team training and team operations.

CHAPTER 1. INTRODUCTION

People can accomplish more work in a team setting than doing work alone (Gorman & Cooke, 2011). This is the case because most work demands more cognitive and/or physical skills and resources than a single person can provide. Increasingly, team-based work has become omnipresent in military and non-military settings (e.g., Salas et al., 2008; Chen, 2018; Gorman et al., 2018). Teaming in the workplace has increased significantly over the past several decades due to changes in the work environment as well as a large number of positive research outcomes from team research (Wynne & Lyons, 2018). Communication is one of the primary factors that researchers focus on when studying teamwork, because it has been consistently linked to team performance outcomes (e.g., Marks et al., 2001; Cooke et al., 2003) and, indeed, has been proposed as an embodiment of team cognition (Cooke et al., 2013).

Traditionally, researchers define teamwork as a process that involves two or more humans working interdependently toward a shared and valued goal (Salas et al., 1992). However, with advances in technology, teamwork has been increasingly extended to teams of human, robots, and artificial intelligence (AI). Therefore, it is necessary to investigate communication patterns not only in human-human teams, but also in human-AI teams to understand differences in human-machine teamwork. The purpose of this study is to investigate whether communication content (what is said) and/or flow (who said it and when) distinguish communication in all-human teams from communication in human-AI teams to learn which aspects of communication predict team performance in all-human versus human-AI teams.

1.1 Team Communication

Team communication can be defined as exchanging information between two or more team members through verbal and/or nonverbal channels (Mesmer-Magnus & DeChurch, 2009) and has been studied as interdependent team behaviors that continuously influence team performance outcomes (Marks et al., 2001). Communication is considered critical to teamwork and coordination because it provides information flows allowing the team to build team cognition, which contributes to situation awareness, decision making, and action at the team level (Cooke et al., 2003). Thus, team communication (verbal communications and nonverbal communications) has been a major focus when conducting teamwork analysis. Foushee and Manos (1981), Orasanu (1990) and Mosier & Chidester (1991), for instance, found that better-performing teams communicated with a higher overall frequency compared to lower performing teams. However, there is also evidence that shows that overall frequency of communication is not the only predictor of team performance. Jentsch and colleagues (1995) discovered that communication content also matters. They found that teams who communicate in more standardized ways, made more leadership statements, and talked more about what they observed in the environment identified a problem significantly faster than those who used fewer of these communication strategies, though they were not faster in solving the identified problem. Though the relationships between communication within teams and team performance depends largely on context (Urban et al., 1995), the studies described above suggest communication as a foundational mechanism that contributes to effective teamwork.

Compared to the somewhat mixed results regarding communication frequency and performance, Marlow and colleagues' meta-analysis (2018) indicates a stronger

relationship between communication quality and team performance. Drawing from various studies analyzing communication quality and/or frequency, Marlow et al., suggest that the act of providing too much information during the communication process is not always beneficial, as it can create unnecessary noise. Furthermore, information elaboration that is defined as the degree to which team members elaborate information shared in the team, was identified as having a stronger relationship with team performance (Homan et al., 2008; Marlow et al., 2018). Therefore, it is useful to take both communication frequency and content quality into account when investigating the relationship between communication and performance in teams.

Communication is an important aspect of team performance in all-human teams, but less is known about teams that consist of human and AI team members. Traditional views on human-centric automation put the human as an authority figure and limited the capabilities of AI (Chen, 2018). The AI can be referred to as automation in this context and requires the human operator's control. Although human-automation interaction has been studied thoroughly, that research may not apply well with the current understanding of the role of AI in teams (Chen, 2020). As technologies advance, researchers have begun to consider AI as fully fledged teammate working with human operators in the team setting (McNeese et al., 2018). Researchers in recent years have argued that it is better to enable dynamic integration of information between human and AI agents through communication that depended on characteristics of the human and AI agent, the context of the task environment, and the task goals (Marathe et al., 2018).

An *autonomous* AI teammate is defined as a type of technology that can collaborate with humans as teammates and perform essential tasks, including making decisions and

executing actions on its own, and has the functions (e.g., communication) required for teamwork (McNeese et al., 2018). In their study, McNeese and colleagues (2018) compared human-AI teams to all-human teams, including control teams that consisted of three novice human operators, and expert teams in which an experimenter served as an expert pilot. The researchers found that although the human-AI teams processed targets in a less efficient manner compared to both the control and expert conditions, because they exhibited different types of communication patterns to process targets (i.e., requesting information more and providing information less compared to all-human teams), overall they performed approximately the same compared to teams in the control condition. This led the researchers to conclude that synthetic AI teammates might replace human teammates for team training and one day might replace human teammates to conduct real-world missions that are difficult or dangerous for humans.

Researchers who propose human-AI teaming as a future direction came to the consensus that bidirectional communication should be implemented, allowing the AI and human team members to understand each other's state (Marathe et al., 2018; Shively et al., 2018). Previous research that focused on human teammates' language preferences in human-robot interaction suggested that from a human teammate's perspective, interacting with autonomy that communicates like a human is preferred (Scalise et al., 2018). However, technical constraints have limited the ability for AI to communicate like humans under many circumstances. Furthermore, it might not always be best to attempt to create an AI that communicates just like a human, because it makes the system overly complex to maintain its function both from a hardware level and from a software level (How, 2016).

At a minimum, the AI should be capable of sending, receiving, and replying to messages from human team members.

1.2 The Current Study

What AI communication characteristics need to be considered for a human-AI team to function most effectively? The current study seeks to answer this question by comparing all-human and human-AI team communication content (actual information shared between team members) and flow (who is communicating to whom and when,). To begin to determine whether communication features distinguish team types (i.e., human-AI vs. all-human team), an initial discriminant analysis on rudimentary communication flow measures from a previous experiment was conducted. Either two human teammates worked with an autonomous AI teammate, or all teammates were human. All teams operated a simulated remotely-piloted aircraft (RPA) to take photos of ground targets. The pilot analysis showed promising results by identifying features of communication flow that distinguished between all-human and human-AI teams (described later in the Data Analysis section). However, that pilot analysis is substantially extended in the current study.

Because the AI's communication capabilities are limited compared to a human operator (e.g., AI is not able to elaborate information sent from human teammates if the message is not written in a format that can be processed by the agent), the communication frequency in human-AI teams is likely less than in all-human teams. In addition, human operators can better elaborate on the information even if the information is not provided in a full sentence; hence, the current study expect all-human teams to process information more efficiently than human-AI teams. Learning from the previous studies regarding

communication frequency and quality, the current study predicts that certain aspects of communication flow and content can serve as classifiers of all-human vs. human-AI team interaction. Therefore, the current study's hypotheses examine whether communication flow and content are significant classifiers that can be used to distinguish between all-human and human-AI teams.

Previous research (McNeese et al., 2018) suggested that human-AI teams perform as well as all-novice human teams, but teams that included an expert team member performed the best. Although human operators are not as restricted in how they share information, they were trained with the same information, which included example messages that a specific role would send to other roles. In the actual missions, members in the all-novice teams might still have used the example messages they learned in the training to complete the mission. However, the communication content in teams that have an expert team member might include more variety. For example, the expert would initiate more positive information such as "thanks" or "roger that". This communication content difference can also lead to communication flow differences, such as promoting closed loop communications in teams with an expert. Therefore, the current study hypothesizes that both communication content and flow can explain the performance differences in McNeese and colleagues' study.

For communication flow, the current study examined measures of communication determinism and transition probabilities between team members. Communication determinism is often used to explore system transitions—the ability for teams to rapidly transition their coordination patterns in response to the changing task environment—which includes changes in team communication flow (Gorman et al., 2012). According to

previous studies (Gorman et al., 2010; Gorman et al., 2012), intact teams exhibit rigid coordination dynamics—reflecting inflexibility in their coordination patterns—and, therefore, generate high determinism. For the current study, although teams in the human-AI condition, the all-novice ("benchmark") condition, and the expert conditions are all intact teams, the flexibility for teams in the human-AI condition is limited because the agent can only recognize communication sequences following specific rules. Therefore, the current study expects teams in the human-AI condition to have higher determinism than teams in the other two conditions.

For transition probabilities, due to the synthetic agent's limited coordination skills, participants must communicate with the synthetic agent following a set of communication guidelines provided to them. Thus, it is expected that there will be less communication when the AI agent serves as a teammate, compared to teams with all human teammates. Furthermore, due to the rather fixed communication happening between human operators and the AI teammate, the current study expects longer "chains" (i.e., significant sequences of communication transitions) in human-AI teams, as their communication are less flexible and should follow more predictable sequences. This prediction is also logically consistent with increased determinism in these teams. Taken together, the current study hypothesizes that communication flow combining communication determinism and transition probabilities should discriminate between all-human and human-AI teams.

Hypothesis 1: Communication flow should discriminate between all-human and human-AI teams.

Hypothesis 2: Teams that include the AI as a teammate have higher determinism than all-human teams.

Hypothesis 3: Less communication occurs when the AI serves as a teammate compared to all-human teams.

Hypothesis 4: Teams that include the AI as a teammate have longer chains compared to all-human teams.

For communication content, because the interaction task that each team in this study uses is relatively fixed - the navigator always needs to send waypoint information to the pilot, the pilot always needs to confirm the settings are sufficient for the photographer to take a good photo, and the photographer always needs to notify the pilot and the navigator that a good photo is taken; the current study expects the communication content in all-human and human-AI teams to be limited to the task itself. In addition, the AI team member works with a restricted vocabulary, which should cause certain communication content (e.g., positive, closed loop communication) to be different between all-human teams and human-AI teams. Thus, the current study hypothesizes that communication content should also discriminate between these team types.

Hypothesis 5: Communication content should discriminate between all-human and human-AI teams. Specifically, due to differences in restricted vs. more open-ended vocabulary, human-AI, all-novice, and expert teams should form unique clusters in terms of semantic similarity.

Finally, based on a previous study suggesting that the performance of expert teams was better than the performance of all-novice and human-AI teams (McNeese et al., 2018), the current study expects similar results in this study.

Hypothesis 6: Communication flow and content of the all-human teams (all-novice; expert) and human-AI teams should predict team performance differences across these conditions.

By testing these hypotheses, the communication feature and pattern analyses in the current study can help us understand whether and how communication differs between human-AI teaming to all human teaming, and how those communication differences are related to team performance.

CHAPTER 2. METHODS

2.1 Participants

For the current study, data were collected at the Cognitive Engineering Research Institute in Mesa, AZ. Seventy graduate and undergraduate students comprising 30 teams were recruited. The 30 teams were divided into three conditions. In the benchmark condition, each team included three naïve participants. In the expert and human-AI conditions, each team included two naïve participants. Participants were required to have a normal or corrected-to-normal vision as well as be fluent in English. Participants aged between 18 to 38 years of age (M = 23.7, SD = 3.3) with a gender distribution of 60 males and 10 females. Each participant was compensated $10 per hour for participating in the study.

2.2 Materials

2.2.1 CERTT Lab

The experiment was conducted in the Cognitive Engineering Research on Team Tasks RPAS Synthetic Task Environment (CERTT-RPAS-STE; Cooke & Shope, 2004). The team's task was to fly a simulated remotely-piloted aircraft (RPA) to take good photos of ground target waypoints. The team includes three distinctive roles: (1) navigator (Data Exploitation, Mission Planning and Communications Operator; DEMPC), who constructs the flying route and provides restriction information to the pilot; (2) pilot (Air Vehicle Operator; AVO), who adjusts the aircraft's altitude, airspeed, and bearing based on target information and restrictions sent from the navigator, and communicates with the

photographer to ensure the RPA settings are adjusted so that the photographer can take a good photo of the target waypoint; and (3) photographer (Payload Operator; PLO), who manages the camera settings based on the information received from the pilot to take a good photograph of each target waypoint. The goal of all teams was to take as many good photographs of strategic target waypoints as possible during a series of 40-minute missions.

Each team member was seated in a separate room at a workstation with monitors, a keyboard, and a mouse. Each workstation had different information on the display based on the team member role (Cooke et al., 2007). Team members communicated with each other through text-based chat. The display for each role and the chat display are illustrated in Figure 1.

All roles were filled by naïve participants in the benchmark condition. In the expert condition, the AVO role was filled by the same confederate in every mission. The confederate had enough experience to be able to communicate and coordinate efficiently with the photographer and the navigator. In the human-AI condition, the AVO role was filled by a synthetic teammate pilot (Ball et al., 2010; McNeese et al., 2018). The synthetic teammate was designed to have a certain level of autonomy in that it could control its action (controlling vehicle altitude and airspeed) and change the action based on information received from text chat during the task. In all conditions, the three team members had to coordinate and share role-specific information to achieve their goal of taking good reconnaissance photos.

Experimenters initiated and ended the experiment and monitored team interaction from a separate workstation, and experimenters were responsible for communicating with

the team in the role of "intelligence" if the team had questions during a mission. The experimenters would stop the mission before the 40-minute time limit if the team reported that they completed all mission goals.

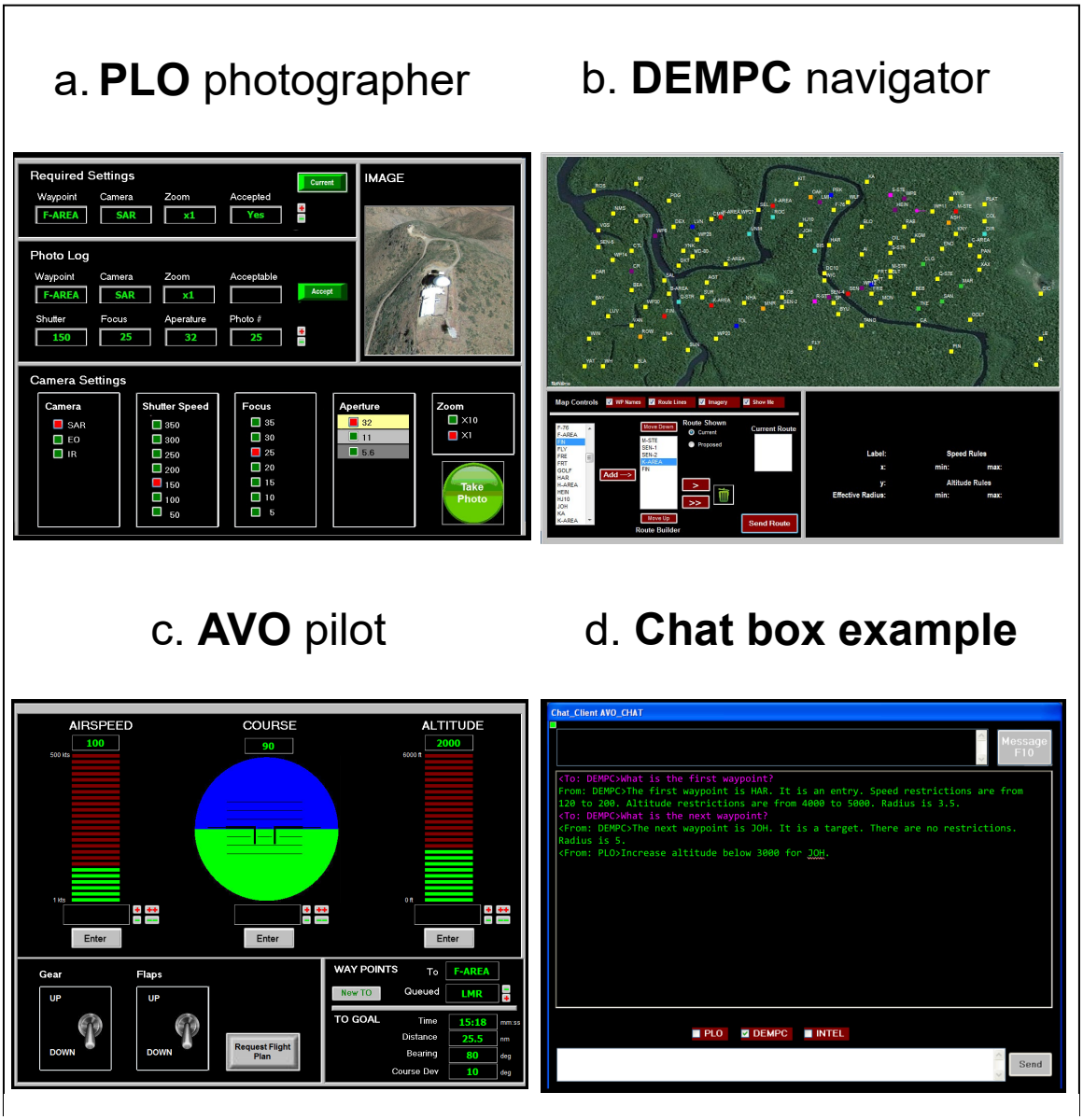

Figure 1 – a. The photographer's workstation display contains waypoints information, camera settings, and photos taken by the photographer; b. The navigator's workstation display contains an area map and waypoint information; c. The pilot's workstation display contains waypoints information, airspeed, altitude, and other flying information; d. Chat interface in the navigator's workstation, which contains recipient selection options and a text box.

2.2.2 The Synthetic Teammate

In the human-AI condition, the pilot role was played by a synthetic team member, which was developed using Adaptive Control of Thought-Rational (ACT-R) cognitive modeling architecture (Anderson, 2007) to simulate AI behavior. The agent is capable of interacting with human teammates via text-based chat. The synthetic agent not only performs the pilot's task but also integrates the text information sent from the other team members into its performance. Similarly, the photographer and the navigator must integrate the information sent from the synthetic agent to perform each mission. Though the synthetic agent has a certain level of autonomy in that it can autonomously carry out its actions as the task unfolds, it is not equal to a human teammate in the sense that it is not explicitly designed to coordinate well with human teammates, especially in a timely manner. Nonetheless, the agent has its own dialog management system that makes the agent capable of managing information requests from other team members and to ask for updates.

Participants in the human-AI condition were told that the pilot was a synthetic agent and were instructed to interact with the synthetic agent using a restricted language (i.e., in a certain format and without any typos). An example of a good communication sequence between the synthetic agent and the human teammates is (1) the navigator sends waypoint information including altitude and airspeed restrictions as well as the effective radius to the synthetic agent; (2) the synthetic agent negotiates a target-specific altitude and airspeed with the photographer; (3) the photographer takes a good photo and sends feedback to both the navigator and the synthetic agent. When the synthetic agent receives feedback from the photographer, it knows that a good photo has been taken and can fly to the next waypoint.

2.3 Experimental Design

The study used a mixed factorial design, with team condition (all-novice teams; expert teams; human-AI teams) being the manipulated between-subjects independent variable and mission number being the within-subject variable. Participants were randomly assigned to the three between-subjects conditions, interacting with either the synthetic agent (human-AI), another participant (all-novice/benchmark), or the trained experimenter as the pilot (expert). All participants completed five 40-minute missions.

2.4 Procedure

Participants were randomly assigned to an RPA team role before they arrived for the experiment. They were assigned to either the pilot (AVO; only in the all-novice/benchmark condition), the navigator (DEMPC), or the photographer (PLO) role. Upon arrival, participants signed an informed consent document and proceeded to a role-specific workstation.

Before the first mission, participants received 30 minutes of PowerPoint training covering general RPA task knowledge, role-specific information, and other roles' responsibilities. After participants completed the PowerPoint training, all participants were given a 30-minute hands-on practice mission familiarizing them with their responsibilities. Participants in the human-AI condition were able to practice communicating with the synthetic teammate during this time. In the expert condition, the confederate used a script to request information if it was not provided in a timely manner. Experimenters provided help for any questions during the training session and ensured that everyone could perform their roles' duties. Participants were then given a 15-minute break.

After training, the experiment began with each team performing five 40-minute missions. The duration of the experimental session was eight hours. The team flew five missions with a 15-minute break after Mission 1, a 30-minute lunch break after Mission 2, and 15-minute breaks after Missions 3, 4, and 5. After all missions were completed, demographic information was collected, and the participants were debriefed.

2.5 Measures

2.5.1 Communication Measures

A chat log recorded the message sent time in seconds, the sender and receiver(s) of the message, when the message was received, and the content of the message. The experimenters could view these interactions from their workstation. During each mission, experimenters monitored the chat and coded the interactions for team process behaviors (not analyzed here). At the end of the experimental session, experimenters saved the chat log output as an Access database document. Communication features for the current study were taken from these Access database files to test which of the features predict team type (human-AI team or all-human team) and how they relate to team performance.

2.5.1.1 Communication Flow

Ordered sequences of chat codes (one for each team and each mission) were generated as the input for the communication flow measures. The input included ordered sequences of mutually exclusive nominal codes for each mission in two ways. The first way was coding the sequence based on sender-receiver differences: 1 = AVO Sending Message to PLO; 2 = AVO Sending Message to DEMPC; 3 = AVO Sending Message to

All; 4 = PLO Sending Message to AVO; 5 = PLO Sending Message to DEMPC; 6 = PLO Sending Message to All; 7 = DEMPC Sending Message to AVO; 8 = DEMPC Sending Message to PLO; 9 = DEMPC Sending Message to All. The second way was coding only the sender sequence: 1 = AVO Sending Message; 2 = DEMPC Sending Message; and 3 = PLO Sending Messages.

2.5.1.1.1 Communication Determinism

The first communication feature that the current study investigated was communication determinism (%DET), which was calculated using discrete recurrence plots (Webber & Zbilut, 1994; Gorman et al., 2012). Discrete team communication states taken from the set of mutually exclusive codes identifying which team member is sending the message and which team member(s) receives it are ordered in a sequence and analyzed using a recurrence plot. Figure 2 shows an example of chat events lined up on the x-axis representing a sequence of codes over time. The recurrence plot takes in this communication sequence and forms a symmetric matrix by lining up the same sequence on the y-axis. Recurrence points are plotted whenever a code at x_i is the same as a code at y_i. The main diagonal in the matrix is trivial, as it contains all of the events plotted against themselves. Recurrence points located off of the main diagonal indicate when codes at one time match codes at earlier and later times. Recurrence points forming diagonals above the main diagonal are called patterns and indicate sequences of speaker codes that repeat over time. These patterns are used to calculate %DET.

%DET is the percentage of the number of recurrent points forming diagonals divided by total recurrent points. Because the recurrence plot is symmetric, only the upper triangle

above the main diagonal is analyzed. %DET ranges from 0 (random/no patterning) to 100 (perfectly repeating pattern). In real-world applications, however, %DET usually lies between these extremes (Gorman et al., 2012). In the current study, for each RPA mission, a discrete recurrence plot was created. %DET scores were then calculated from recurrent diagonals relative to all recurrent points, using Equation 1:

$$\%\text{DET} = \frac{\text{number of recurrent points forming diagonals}}{\text{total number of recurrent points}} \text{ X } 100 \qquad (1)$$

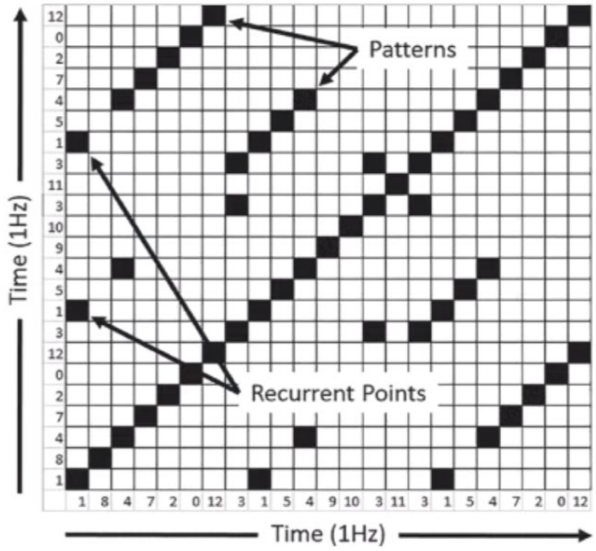

Figure 2 – Example discrete recurrence plot from Gorman et al., 2020.

2.5.1.1.2 Transition Probabilities

Transitions between senders and receivers as well as just senders were calculated to obtain the transition probabilities of each sender-receiver combination (AVO to PLO, AVO to DEMPC; PLO to AVO, PLO to DEMPC; DEMPC to AVO, DEMPC to PLO), as well

as the transition probabilities of sender sequences. First, a raw transition matrix was formed for each mission (e.g., Table 1), and transition probabilities were then be computed by dividing each transition frequency by the row frequency as a separate communication feature for the analyses (e.g., Table 2). A transition probability is a type of conditional probability that measures the observed probability (relative frequency) of all "from-to" combinations over a mission. Simple transition probabilities ("Lag-0") measure transition probabilities from time t to time $t + 1$, as shown in Table 2. To identify longer from-to sequences ("chains"), the lag is increased (e.g., Lag-1; Lag-2; etc.). As represented as event sequences, Lag 0 represents the probability of observing event B given event A, right after event A; Lag 1 represents the probability of observing event B given event A, after one intervening event, and so on (Bakeman & Gottman, 1997). Based on prior research (Kiekel et al., 2002), transition probabilities are computed up to four lags in the current study.

After obtaining the lagged transition probabilities, to calculate which transition probabilities occur significantly greater than chance, a *z*-score approach for Lag-sequential modeling was conducted (O'Connor, 1999). Z scores are called adjusted residuals in the Lag-sequential modeling, and these scores are assessed by referencing to the standard normal distribution. Z scores are used to identify the significance of transition probabilities (O'Connor, 1999). The current study used Lag-sequential modeling to look for statistically significant higher-order lags of transition probabilities (i.e., significant chains of communication events; e.g., AVO→PLO→DEM would be a lag two chain). Max chain length is another communication feature that measures the consistency of communication pattern in teams. The max chain length was determined based on summing up the frequency of significant transition probabilities up to lag four in a single mission. If there was a

significant transition probability within a specific lag, the frequency would be "1" for that lag. An example of max chain length calculated for a mission is illustrated in Table 3.

Table 1 – Transition Matrix.

		To		
		AVO	DEMPC	PLO
From	AVO	0	23	13
	DEMPC	30	0	1
	PLO	16	1	13

Table 2 – Transition Probability Matrix.

		To		
		AVO	DEMPC	PLO
From	AVO	0	0.6389	0.3611
	DEMPC	0.9677	0	0.0323
	PLO	0.9412	0.0588	0

Table 3 – Max Chain Length.

Team	Mission	Lag 1	Lag 2	Lag 3	Lag 4	Max Chain Length
1	1	1	1	1	0	3

2.5.1.2 Communication Content

Content information—what was said—is also recorded in the chat log output. To measure communication content features, the current study performed Latent Semantic Analysis (LSA).

2.5.1.2.1 LSA

LSA is a mathematical/statistical method used to represent and analyze semantic knowledge in a domain of discourse. It was developed based on the theory that knowledge is reflected in how word meaning derives from the surrounding context of words within meaningful discourse (Landauer & Dumais, 1997). Previous research has concluded that the main advantage of LSA over other communication content analysis tools is that LSA is examining the semantic relatedness of utterances of the conversation rather than focusing on the meanings of individual words (Dong, A. 2005).

LSA compares team communication (e.g., chat messages) to a semantic space, where the semantic space represents a factor analytic model of the domain of discourse (e.g., Gorman et al., 2016). The current study used the topic space named "General Reading up to 1st year college" provided on the LSA Colorado website (lsa.colorado.edu) as well as a custom UAV semantic space created for the current study.

To create a semantic space, LSA converts the model input (called a "corpus" or a body of text: e.g., manuals and transcripts) to a frequency co-occurrence matrix of terms (unique words; rows) by documents (paragraphs; columns). The UAV semantic space created for this study contains inputs from a corpus that included the Unmanned Aerial

Vehicle Aircrew Training Manual and the Remote Pilot Study Guide, in addition to the one-hundred-and-fifty mission transcripts from the current dataset (previous research identified including transcripts in the corpus as a standard practice; e.g., Gorman et al., 2016). The co-occurrence matrix dimensions were 5,087 unique words × 1,688 unique paragraphs.

LSA then reduces the dimensionality of the frequency co-occurrence matrix to compute the underlying, latent semantic factors using singular value decomposition (SVD). SVD is similar to principal components/factor analysis in that the magnitude of singular values corresponds to how salient the factors are but can be performed on a rectangular matrix. The optimal number of dimensions (factors) was determined based on results and recommendations from previous studies (Landauer et al., 1998; Gorman et al., 2016). There were 300 dimensions in the UAV semantic space.

Two LSA metrics used in the current study are (1) the vector length of a piece of discourse and (2) the cosine between two pieces of discourse (cosine similarity). The vector length measures the amount of speech weighted by the discourse's domain-specific content. The cosine is the dot product between two pieces of discourse plotted in the semantic space, which measures the correlation between any two pieces of discourse regardless of the utterances' length or the time the utterances occur (e.g., Gorman et al., 2016).

These two metrics were used to analyze the amount of domain-relevant knowledge contained on average in the chat messages and the average semantic relatedness of chat messages in terms of RPA-relevant content during each mission. Due to computational

difficulties, the "General Reading up to 1st year college" topic space (lsa.colorado.edu) was used to compute vector length. However, the UAV space was used for all cosine similarities analyses, which are the most commonly reported type of LSA analyses.

2.5.2 Team Performance Measures

2.5.2.1 Team Performance Outcome

Team performance was calculated for every mission as the weighted composite of several system parameters, including duration of warning or alarm state, rate of good photographs per minute, fuel and film used, and the number of missed targets. At the beginning of each mission, each team had an initial score of 1,000, and points were deducted based on the final value of each system parameter (Cooke et al., 2007). The current study used the team communication features to predict these team performance scores across the human-AI and all-human conditions.

2.5.2.2 Target Processing Efficiency (TPE)

Target processing efficiency was calculated for every target based on the time spent within a target waypoint's effective radius to get a good photo. Higher TPE scores indicate greater efficiency. For each target, teams had an initial score of 1,000. Points were deducted based on the number of seconds in the effective radius, and an additional 200 points would be deducted if the team failed to get a photo for that target (Cooke et al., 2007). TPE measures teams' efficiency regarding targets and is thus sampled more frequently than team performance outcome, which is based on the overall mission. Similar to team

performance outcome, the current study used the team communication features to predict TPE scores, averaged across mission, for the human-AI and all-human teams.

CHAPTER 3. RESULTS

3.1 H1: Communication Flow will Discriminate Team Type.

To test H1 that communication flow will discriminate between human-AI and all-human teams, the current study performed a discriminant function analysis to check whether the flow features (%DET scores, lag 0, lag 1, and lag 2 transition probabilities for all missions) were included in the discriminant function as predictors. The discriminant function with these flow features as predictors to differentiate human-AI teams from all-human teams was significant (Wilks's $\Lambda = .407$, $\chi^2(18) = 124.837$, $p < .001$). The canonical correlation was .77 for the discriminant functions, indicating that the relationship between selected flow features and team type (human-AI or all-human team) was significant in the discriminant function.

Looking at the standardized canonical discriminant function coefficients (Table 4), the discriminant function had the largest relationship with lag 2 transition probabilities, starting with a DEMPC event, followed by several lag 0, lag 1, and lag 2 transition probabilities started with either AVO, DEMPC, or PLO events, and finally, %DET. However, when looking at the structure matrix (Table 5), the results showed that only lag 1 transition probabilities started with an AVO event and followed by an AVO event, lag 2 transition probabilities started with an AVO event to an AVO event, lag 1 transition probabilities started with a PLO event followed by a DEMPC event, and lag 2 transition probabilities started with a DEMPC event to an AVO event had values above .30 (Brown & Wicker, 2000). This indicated that these four communication flow features were significant predictors in the discriminant function and all other features were considered as

poor predictors. The discriminant function equation bellow presented how each feature contributed to the discriminant function:

Discriminant score = (12.371*DEMPC2AVO) + (14.41*DEMPC2PLO) + (11.365* DEMPC2DEMPC) + (-7.888* PLO1PLO) + (-6.295* PLO2DEMPC) + (-7.185* AVO1AVO) + (-5.544* AVO2AVO) + (5.818* DEMPC1DEMPC) + (3.284* DEMPC1AVO) + (-3.502* PLO2AVO) + (-2.337* PLO1AVO) + (-4.871* AVO2DEMPC) + (-2.519* AVO0PLO) + (1.718* DEMPC0AVO) + (-1.773* PLO1DEMPC) + (1.272* PLO0AVO) + (-1.08* AVO1DEMPC) + (-.001*%DET) – 3.169.

Their group centroids were -1.694 and .847, respectively. Thus, all-human teams scored higher on the discriminant function with selected flow features as predictors than human-AI teams. The cross validated classification showed that human-AI teams and all-human teams were correctly discriminated based on that difference in 88.0% of the cases. Discriminant analysis does not precisely indicate how these predictors differ between human-AI and all-human teams. However, results relevant to this question are included in Hypotheses 2-4. Overall, the results supported the hypothesis, suggesting that communication flow features discriminate between human-AI and all-human teams.

Table 4 – Standardized Canonical Discriminant Function Coefficients for predictors.

Rank	Predictor	Coefficients
1	DEMPC2AVO	1.937
2	DEMPC2PLO	1.891
3	DEMPC2DEMPC	1.396

Table 4 continued.

4	PLO1PLO	-.987
5	PLO2DEMPC	-.919
6	AVO1AVO	-.814
7	AVO2AVO	-.754
8	DEMPC1DEMPC	.687
9	DEMPC1AVO	.681
10	PLO2AVO	-.607
11	PLO1AVO	-.492
12	AVO2DEMPC	-.457
13	AVO0PLO	-.337
14	DEMPC0AVO	.27
15	PLO1DEMPC	-.251
16	PLO0AVO	.213
17	AVO1DEMPC	-.111
18	%DET	-.003

Table 5 – Structure Matrix.

Rank	Predictor	Coefficients
1	**AVO1AVO**	**-0.607**
2	**AVO2AVO**	**-0.361**
3	**PLO1DEMPC**	**0.358**
4	**DEMPC2AVO**	**-0.351**
5	PLO1AVO	-0.238
6	PLO2DEMPC	-0.233
7	DEMPC2PLO	0.225
8	AVO2DEMPC	0.222
9	DEMPC2DEMPC	0.202
10	PLO0AVO	0.173
11	DEMPC1DEMPC	0.135
12	DEMPC1AVO	-0.133

Table 5 continued.

13	PLO2AVO	-0.052
14	DEMPC0AVO	0.037
15	AVO0PLO	0.033
16	%DET	-0.023
17	AVO1DEMPC	-0.005
18	PLO1PLO	0.004

Note. Bolded predictors that were above the .30 cut-off.

3.2 H2: Human-AI Teams have Higher Determinism than All-Human Teams.

To test H2, the current study carried out an independent-samples *t*-test (Table 6) comparing the %DET per mission between human-AI teams and the all-human teams. Results showed that the effect of team type was not significant, $t(148) = -.33$, $p = .739$, $d = -.058$; indicating that all-human teams' ($M = 58.59$, $SD = 3.87$) communication determinism was not statistically different from human-AI teams' ($M = 58.83$, $SD = 4.68$).

Table 6 – Summary of T-Test comparing %DET between Human-AI Teams and All-Human Teams.

	t	df	p	d
Total	-.334	148	.739	-.58

Note. This model tests if the mean %DET in *All-Human Teams* is greater than that in *Human-AI Teams*.

Nonetheless, %DET was included as a predictor in the discriminant analysis. To understand if there were variation based on condition, the current study conducted a one-way ANOVA to compare the effect of condition (Human-AI, Benchmark, and Expert) on %DET. The ANOVA results are summarized in Table 7 and the *post hoc* analysis results

are summarized in Table 8. Results revealed that there was a statistically significant difference in mean %DET score between at least two groups ($F(2, 147) = 6.41$, $p = .002$, $\eta_p^2 = .080$). The Tukey *post hoc* test for multiple comparisons found, surprisingly, that the mean value %DET scores was not significantly different between the human-AI teams ($M = 58.83$, $SD = 4.68$) and all-novice teams ($M = 57.16$, $SD = 4.36$; $p = .097$) nor the human-AI teams and expert teams ($M = 60.013$, $SD = 2.68$; $p = .303$). However, there was a statistically significant difference in mean value of %DET between all-novice teams and expert teams ($p = .001$, 95% C.I. = [-4.75, -.957]).

Together, these results do not support Hypothesis 2, indicating human-AI teams and all-human teams did not differ on communication determinism. While all-novice teams had lower mean %DET score compared to human-AI teams, expert teams had the highest mean %DET value.

Table 7 – Summary of ANOVA Results comparing %DET among Human-AI Teams and All-Novice Teams, and Expert Teams.

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Condition	205.272	2	102.636	6.409	.002	.08
Residuals	2354.079	147	16.014			

Note. This model tests if the mean %DET varies among *Human-AI Teams*, *All-Novice Teams*, and *Expert Teams*.

Table 8 – Summary of Post Hoc Comparisons on Condition.

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
Human-AI	Novice	1.667	-.228	3.562	.8	2.083	.097
Residuals	Expert	-1.185	-3.08	0.71	.8	-1.481	.303
Novice	Expert	-2.852	-4.747	-0.957	.8	-3.563	.001

Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the tukey method).

3.3 H3: Less Communication will Occur in Human-AI Teams compared to All-Human Teams.

To test H3, the current study carried out an independent-samples *t*-test (Table 9) comparing the total frequency of lag 1 transition probabilities per mission between human-AI teams and all-human teams. Results showed a significant effect of team type, $t(148) = 1.97$, $p = .026$, $d = .34$; indicating that all-human teams ($M = 90.16$, $SD = 25.13$) had significantly more communication during the missions compared to human-AI teams ($M = 73.76$, $SD = 75.67$).

Table 9 – Summary of T-Test comparing Communication Frequency between Human-AI Teams and All-Human Teams.

	t	df	p	d
Total	1.967	148	.026	.341

Note. This model tests if the mean lag 1 frequency in *All-Human Teams* is greater than that in *Human-AI Teams*.

To further understand what contributed to this difference, the current study conducted a one-way ANOVA to compare the effect of condition (Human-AI, All-novice, and Expert) on communication frequency. The ANOVA results are summarized in Table 10 and the *post hoc* analysis results are summarized in Table 11. Results revealed that there was a statistically significant difference in mean frequency of lag 1 transition probabilities between at least two groups ($F(2, 147) = 6.92$, $p = .001$, $\eta_p^2 = .086$). The Tukey *post hoc* test for multiple comparisons found that the mean value of lag 1 transition probabilities was not significantly different between human-AI teams and all-novice teams ($p = .980$). However, there were statistically significant differences in mean frequency of lag 1 transition probabilities between human-AI teams and expert teams ($p < .01$, 95% C.I. = [-53.15, -8.85]) and between all-novice teams and expert teams ($p < .01$, 95% C.I. = [-51.35, -7.05]).

Together, these results supported Hypothesis 3, indicating less communication happened when the AI served as a teammate compared to all-human teams. However, this difference was primarily due to expert teams that communicated significantly more than either human-AI teams or all-novice teams.

Table 10 – Summary of ANOVA Results comparing Communication Frequency among Human-AI Teams and All-Novice Teams, and Expert Teams.

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Condition	30281.333	2	15140.667	6.917	.001	.086
Residuals	321758.560	147	2188.834			

Note. This model tests if the mean lag frequency varies among *Human-AI Teams*, *All-Novice Teams*, and *Expert Teams*.

Table 11 – Summary of Post Hoc Comparisons on Condition.

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
Human-AI	Novice	-1.800	-23.954	20.354	9.357	-.192	.980
Residuals	Expert	-31.000	-53.154	-8.846	9.357	-3.313	.003
Novice	Expert	-29.200	-51.354	-7.046	9.357	-3.121	.006

Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the tukey method).

3.4 H4: Longer Chains in Human-AI Teams compared to All-Human Teams.

To test H4, the current study compared the *z* scores for lag 1 and lag 2 transition probabilities in human-AI teams and all-human teams to determine if human-AI teams have more lag one and lag two chains than All-Human Teams. The results are summarized in Table 12.

Based on the results, the lag one chains that human-AI teams had more than all-human teams were AVO → AVO, and AVO → DEMPC. The lag two chains that human-AI teams had more than all-human teams was DEMPC → → AVO. The lag one chains that all-human teams had more than human-AI teams were AVO → PLO, and PLO → DEMPC. The lag two chains that all-human teams had more than human-AI teams were AVO → → DEMPC and PLO → → AVO. The results indicated that for just lag one and lag two chains, human-AI teams did not have more or longer chains than all-human teams.

Table 12 – Summary of T-Test comparing Z Scores for Lag 1&2 Transition Probabilities between Human-AI Teams and All-Human Teams.

Feature	t	df	p
AVO1AVO	-3.889	148	$< .001^{*}$
AVO1DEMPC	-2.281	148	$.024^{*}$
AVO1PLO	6.824	148	$< .001^{*}$
DEMPC1AVO	2.870	148	$.005^{a}$
DEMPC1DEMPC	-1.758	148	.081
DEMPC1PLO	-1.886	148	.061
PLO1AVO	1.480	148	.141
PLO1DEMPC	4.133	148	$< .001^{*}$
PLO1PLO	-6.090	148	$< .001^{a}$
AVO2AVO	-1.453	148	.148
AVO2DEMPC	2.442	148	$.016^{*}$
AVO2PLO	-1.128	148	.261
DEMPC2AVO	-2.255	148	$.026^{*}$
DEMPC2DEMPC	1.618	148	.108
DEMPC2PLO	.763	148	.447
PLO2AVO	3.201	148	$.002^{*}$
PLO2DEMPC	-3.822	148	$< .001^{a}$
PLO2PLO	.974	148	.332

Note. These model test if the mean z score for lag 1 and lag 2 transition probabilities in *Human-AI Teams* is greater than that in *All-Human Teams*. * Significant results. ^{a} Leven's test is significant ($p < .05$).

Nonetheless, the results did not rule out the possibility of longer lag chains, so the current study carried out an independent-samples *t*-test (Table 12) comparing the max number of chains (summed from lag one to up to lag four transition probabilities) per mission between human-AI teams and the all-human teams. Levene's test was significant, indicating a violation of the equal variance assumption.

Table 13 – Summary of T-Test comparing Max Chain Length between Human-AI Teams and All-Human Teams.

	t	df	p	d
Total	-1.70	148	$.091^{a}$	-.294

Note. This model tests if the mean max chain length in *Human-AI Teams* is differ than that in *All-Human Teams*.

To further investigate whether the max chain length differs among the three conditions, the current study performed a one-way ANOVA to compare the effect of condition (Human-AI, Benchmark, and Expert) on the max chain length. The ANOVA results are summarized in Table 14. Results revealed that there was no significant difference in mean max chain length between at least two groups ($F(2, 147) = 2.47$, $p = .088$, $\eta_p^2 = .032$).

Together, these results do not support Hypothesis 4, indicating human-AI teams and all-human teams did not differ on chain length.

Table 14 – Summary of ANOVA Results comparing Max Chain Length among Human-AI Teams and All-Novice Teams, and Expert Teams.

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Condition	4.120	2	2.060	2.466	.088	.032
Residuals	122.820	147	.836			

Note. This model tests if the mean max chain length varies among *Human-AI Teams*, *All-Novice Teams*, and *Expert Teams*.

3.5 H5: Communication Content will Discriminate Team Membership.

3.5.1 Vector Length

To test H5, the current study first performed an independent-samples *t*-test (Table 15) comparing the averaged vector length per mission between human-AI teams and all-human teams. The Levene's test was significant indicating a violation of the equal variance assumption.

Table 15 – Summary of T-Test comparing Vector Length between Human-AI Teams and All-Human Teams.

	t	df	p	d
Total	1.259	148	$.210^{a}$	.218

Note. This model tests if the mean vector length in *All-Human Teams* is greater than that in *Human-AI Teams*. ^{a} Leven's test is significant ($p < .05$).

To further investigate whether vector length differs among the three conditions, the current study performed a one-way ANOVA to compare the effect of condition (Human-AI, Benchmark, and Expert) on vector length. The ANOVA results are summarized in Table 16 and the *post hoc* analysis results are summarized in Table 17. Results revealed that there was a statistically significant difference in mean vector length between at least two groups ($F(2, 147) = 8.75$, $p < .001$, $\eta_p^2 = .106$). The Tukey *post hoc* test for multiple comparisons found that the mean value of vector length was not significantly different between human-AI teams and all-novice teams ($p = .678$). However, there were statistically significant differences in mean vector length between human-AI teams and expert teams ($p < .01$, 95% C.I. = [-.571, -.079]) and between all-novice teams and expert teams ($p < .001$, 95% C.I. = [-.658, -.166]). The results indicated that expert teams had more domain-specific content of utterances compared to all-novice and human-AI teams.

Table 16 – Summary of ANOVA Results comparing Vector Length among Human-AI Teams, All-Novice Teams, and Expert Teams.

Cases	Sum of Squares	df	Mean Square	F	p	η^2_p
Condition	4.717	2	2.359	8.749	< .001	.106
Residuals	39.630	147	.270			

Note. This model tests if the mean vector length among *Human-AI Teams*, *All-Novice Teams*, and *Expert Teams*.

Table 17 – Summary of Post Hoc Comparisons on Condition.

			95% CI for Mean Difference				
		Mean Difference	Lower	Upper	SE	t	p_{tukey}
Human-AI	Novice	.087	-.158	.333	.104	.842	.678
Residuals	Expert	-.325	-.571	-.079	.104	-3.128	.006
Novice	Expert	-.412	-.658	-.166	.104	-3.969	< .001

Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the tukey method).

3.5.2 K-Means Cluster

For the cosine similarities (semantic relatedness) feature, the current study conducted K-Means clustering of the mission × mission cosine matrix. This matrix contained the cosines (essentially, correlations) between all mission transcripts across all experimental conditions (150 × 150). To determine the optimal number of clusters, the current study computed the K-Means elbow function. The result indicated that 3 was the optimal number for K-Means clustering (Figure 3).

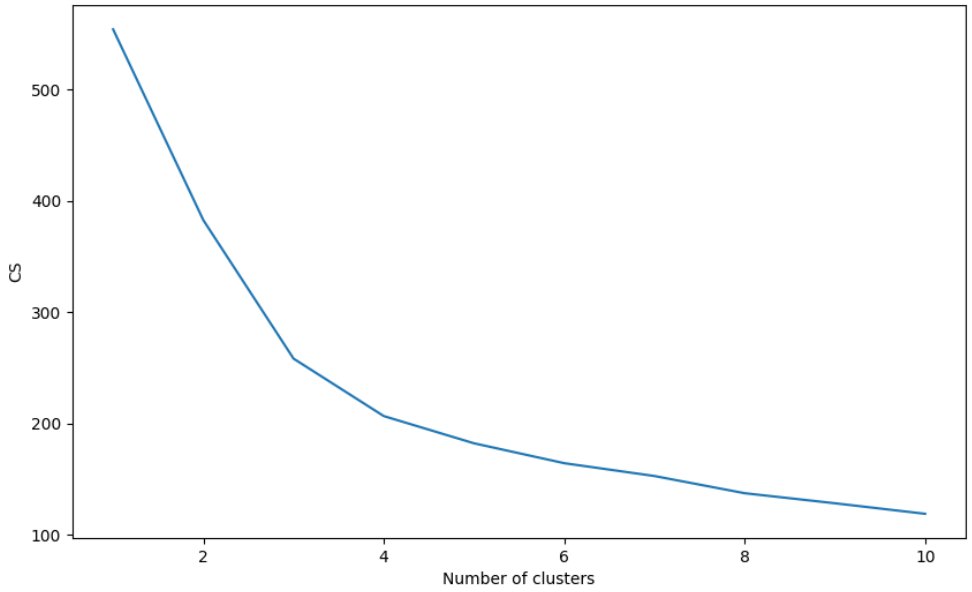

Figure 3 – K-Means Elbow Function Result.

The K-Means clustering results are presented in Figure 4. K-Means cluster centroids were calculated, and the Euclidean distances among the three clusters are summarized in Table 18. The Euclidean distances indicate the degree of similarities between the cluster's centroids.

To examine K-Means clustering classification results, the clustered were relabeled based on condition. As shown on the Figure 5, expert teams and human-AI teams each clustered based on communication differences, whereas all-novice teams were more dispersed. Overall, K-Means clustering successfully classified 98% of the human-AI teams, 88% of the expert teams, but only 50% of the all-novice teams (Table 19).

Table 18 – Euclidean Distances of Cluster Centroids.

Cluster - Condition	Cluster		Euclidean distance
0 – Expert	0	1	2.83
1 – Human-AI	0	2	2.08
2 – All-Novice	1	2	3.26

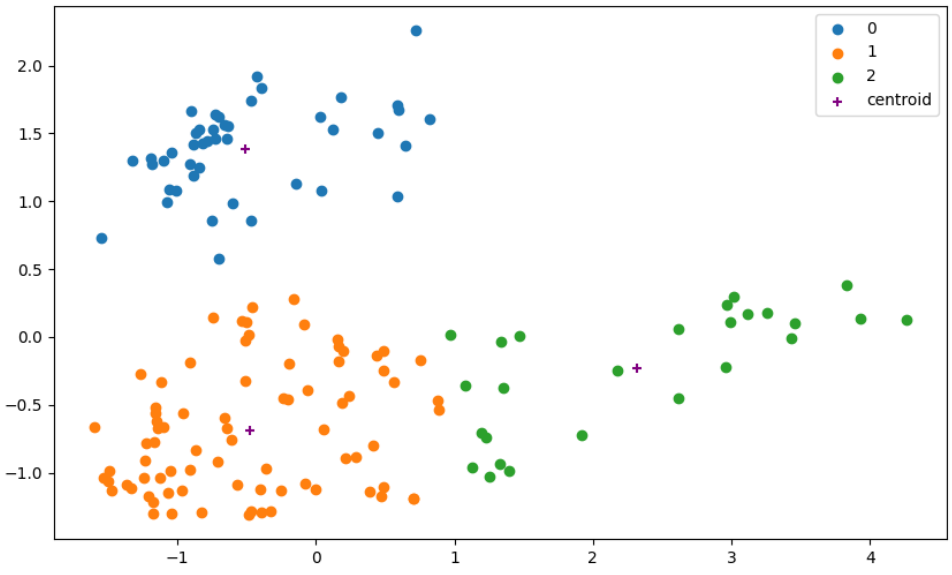

Figure 4 – K-Means Clustering Results.

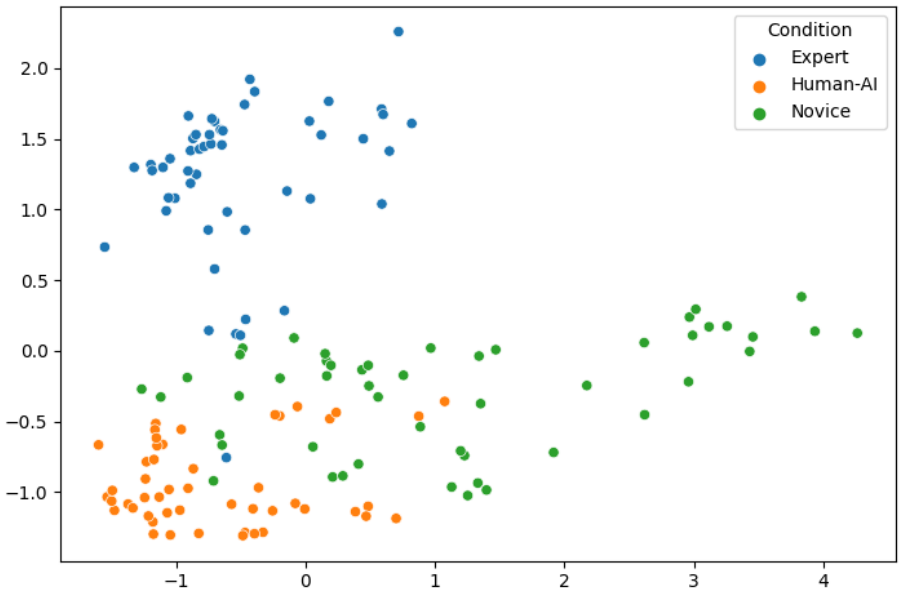

Figure 5 – K-Means Clusters Relabeled Based on Condition.

Table 19 – Summary of K-Means Clustering Classification Results.

	Predicted	Total	%Correctness
Human-AI	49	50	98%
Novice	25	50	50%
Expert	44	50	88%

3.5.3 Discriminant Function Analysis

To further examine whether communication content features can discriminate team membership, the current study conducted a discriminant analysis to predict whether the team was human-AI team or all-human team with the vector length and K-Means results

of the mission cosine matrix as predictors. The discriminant function with just the content features as predictors to differentiate human-AI teams from all-human teams was significant (Wilks's $\Lambda = .643$, $\chi^2(2) = 64.913$, $p < .001$). The canonical correlation was .597 for the discriminant function. The discriminant function had a larger relationship with K-Means of the cosine matrix (1.011), followed by vector length (-.066). The structure matrix corresponds to the relationships, indicating K-Means of the cosine matrix as a significant predictor in the discriminant function (.998) and vector length as a poor predictor. Their group centroids were .523 and -1.047, respectively. Hence, all-human teams scored higher on the content discriminant function than human-AI teams. Human-AI teams and all-human teams were correctly discriminated based on that difference in 78.7% of the cases. The discriminant function is:

Discriminant score = (1.427*K-Means) + (-.122*vector length) -.908.

Furthermore, the current study added the vector length and K-Means results of the mission cosine matrix into the discriminant function input and performed discriminant function analysis using both flow and content features. The results showed an increase in the discriminant function predictability.

Overall, the discriminant function with flow and content features as predictors to differentiate human-AI teams from all-human teams was significant (Wilks's $\Lambda = .264$, $\chi^2(20) = 183.613$, $p < .001$). The canonical correlation was .858 for the discriminant functions, indicating that 85.8% of variance was explained by the relationship between selected flow and content features and team type (human-AI team or all-human team) in the discriminant function.

There were some changes with the relationships between discriminant function and the predictors. Specifically, vector length and K-Means of cosine similarity matrix were included in the discriminant function. Although the discriminant function still had the largest relationship with lag 2 transition probabilities started by a DEMPC event, there were slight differences on the lag 0, lag 1, and lag 2 transition probabilities started with either AVO, DEMPC, or PLO events compared to the earlier analysis. The standardized canonical discriminant function coefficients for all identified predictors are reported in Table 20. Taking a closer look at the structure matrix (Table 21), adding K-Means of cosine similarity matrix and vector length resulted in decreases in the original flow predictors' loadings. Only K-Means of cosine similarity matrix and lag 1 transition probabilities started with an AVO event followed by an AVO event were significant (coefficients above .30). The discriminant function becomes:

Discriminant score = (15.752*DEMPC2AVO) + (17.063*DEMPC2PLO) + (15.206* DEMPC2DEMPC) + (1.258*K-Means) + (-7.204* AVO1AVO) + (-5.087* AVO2AVO) + (3.169*PLO1AVO) + (3.024*DEMPC0AVO) + (3.027*PLO1DEMPC) + (-2.916*AVO0PLO) + (-2.592*PLO2DEMPC) + (-1.911*PLO2AVO) + (-2.873*AVO1DEMPC) + (2.418*DEMPC1DEMPC) + (-2.596*AVO2DEMPC) + (.561*DEMPC1AVO) + (.205* Vector Length) + (-.489* PLO0AVO) + (.562*PLO1PLO) + (.011*%DET) – 12.703.

All-human and human-AI teams' group centroids were 1.172 and -2.343, respectively. Hence, all-human teams scored higher on the combined flow and content discriminant function than human-AI teams. Human-AI teams and all-human teams were correctly discriminated based on that difference in 91.3% of the cases, which is a 3.3%

increase compared to only including flow features in the discriminant function. To conclude, the results supported H5, suggesting that cosine similarities (semantic relatedness) are communication content features that discriminate between human-AI teams and All-Human Teams.

Table 20 – Standardized Canonical Discriminant Function Coefficients for predictors after adding Communication Content.

Rank	Predictor	Coefficients
1	DEMPC2AVO	2.466
2	DEMPC2PLO	2.239
3	DEMPC2DEMPC	1.868
4	**K-Means**	**.891**
5	AVO1AVO	-.817
6	AVO2AVO	-.692
7	PLO1AVO	.667
8	DEMPC0AVO	.476
9	PLO1DEMPC	.429
10	AVO0PLO	-.39
11	PLO2DEMPC	-.378
12	PLO2AVO	-.331
13	AVO1DEMPC	-.295
14	DEMPC1DEMPC	.285
15	AVO2DEMPC	-.244
16	DEMPC1AVO	.116
17	**Vector Length**	**.111**
18	PLO0AVO	-.082
19	PLO1PLO	.07
20	%DET	-.046

Note. Bolded predictors that were communication content features.

Table 21 – Structure Matrix after adding Communication Content.

Rank	Predictor	Coefficients
1	**K-Means**	**0.446**
2	AVO1AVO	-0.439
3	AVO2AVO	-0.261
4	PLO1DEMPC	0.259
5	DEMPC2AVO	-0.254
6	PLO1AVO	-0.172
7	PLO2DEMPC	-0.168
8	DEMPC2PLO	0.163
9	AVO2DEMPC	0.16
10	DEMPC2DEMPC	0.146
11	PLO0AVO	0.125
12	DEMPC1DEMPC	0.098
13	DEMPC1AVO	-0.096
14	**Vector Length**	**0.062**
15	PLO2AVO	-0.038
16	DEMPC0AVO	0.027
17	AVO0PLO	0.024
18	%DET	-0.016
19	AVO1DEMPC	-0.003
20	PLO1PLO	0.003

Note. Bolded predictors that were communication content features.

3.6 H6: Communication Flow and Content Predicting Human-AI Team and All-Human Team Performance.

3.6.1 Correlation Analyses

The sixth hypothesis tested whether the selected communication flow and content features can predict team performance variation across conditions. To test H6, zero-order Pearson correlations were performed to investigate the relationships between the

communication flow and content features selected from the discriminant function analysis and the outcome measures (team performance and TPE; Table 22).

Table 22 – Summary of Pearson Correlation Coefficients.

Features	Team Performance Score	Average TPE Score
AVO0PLO	r = -.029, p = .727	r = .098, p = .232
PLO0AVO	r = .088, p = .284	r = .203, p = .013
DEMPC0AVO	r = .19, p = .02	r = .102, p = .215
AVO1AVO	r = -.107, p = .193	r = -.243, p = .003*
AVO1DEMPC	r = -.072, p = .378	r = .03, p = .715
DEMPC1AVO	r = .324, p < .001**	r = .209, p = .01
DEMPC1DEMPC	r = -.279, p < .001**	r = -.08, p = .33
PLO1AVO	r = .202, p = .013	r = .113, p = .167
PLO1DEMPC	r = -.042, p = .612	r = .074, p = .369
PLO1PLO	r = -.306, p < .001**	r = -.302, p < .001**
AVO2AVO	r = .127, p = .12	r = -.045, p = .586
AVO2DEMPC	r = -.142, p = .084	r = -.007, p = .93
DEMPC2AVO	r = .097, p = .238	r = -.011, p = .891
DEMPC2DEMPC	r = -.152, p = .064	r = -.067, p = .319
DEMPC2PLO	r = .014, p = .867	r = .069, p = .403
PLO2AVO	r = .191, p = .019	r = .147, p = .072
PLO2DEMPC	r = -.165, p = .044	r = -.097, p = .236
%DET	r = .098, p = .231	r = .032, p = .696
Vector Length	r = .2, p = .014	r = .17, p = .037
K-Means	**r = .477, p = < .001****	**r = 0.571, p = < .001****

Note. Correlations of communication flow and content features with outcome measures. Medium to large correlations are in bold, with asterisks denoting the following *p < .01, **p < .001.

Based on these correlation results, the current study found correlations of medium sizes for K-Means group membership based on cosine similarities, indicating an association of communication content with both team performance (r = .477, p = < .001) and TPE (r = 0.571, p = < .001). Due to the large number of correlations, a medium-to-

large effect size criteria was used to identify meaningful correlations. Although there are several other significant correlations, many of the results do not fulfill the medium-to-large effect size criteria. Nonetheless, the correlation analysis only looks at relationships between communication features and outcome variables at the individual level, while the discriminant function combines these features to perform classification. Therefore, the current study also examined a combination effect of communication flow and content features on outcome variables.

3.6.2 Multiple Regression

To further examine H6, the current study performed a multiple regression to investigate whether the selected communication flow and content features could significantly predict outcome variables (Team performance, TPE).

3.6.2.1 Team Performance

The results of the regression indicated that the model explained 43% of the variance and that the model was a significant predictor of team performance, $F(20,129) = 4.88$, $p < .001$. The coefficients are summarized in Table 23. Based on the results, the lag 1 transition probabilities with the sequence AVO → AVO (B = -521.93, β = -.527, t = -2.57, p =.011), DEMPC → DEMPC (B = -387.51, β = -.333, t = -2.17, p =.032) and K-Means of the cosine similarities (B = 63.223, β = .166, t = 4.58, $p < .001$) contributed significantly to the model. The final predictive model was:

Team performance score = 604.924 + (-1.124*AVO0PLO) + (-124.18*PLO0AVO) + (80.283*DEMPC0AVO) + (-521.93*AVO1AVO) + (-75.135* AVO1DEMPC) + (41.348* DEMPC1AVO) + (-387.51* DEMPC1DEMPC) + (221.612* PLO1AVO) + (228.966* PLO1DEMPC) + (147.318*PLO1PLO) + (201.061* AVO2AVO) + (74.114* AVO2DEMPC) + (-239.96* DEMPC2AVO) + (-307.89* DEMPC2DEMPC) + (-303.98*

DEMPC2PLO) + (132.727* PLO2AVO) + (75.22* PLO2DEMPC) + (-4.729*%DET) + (10.269* Vector Length) + (63.223* K-Means)

Table 23 – Summary of Multiple Regression Analysis for Team Performance.

	B	*SE B*	*β*	*t*	*p*
(Constant)	604.924	946.643		.639	.524
AVO0PLO	-1.124	93.827	-.001	-.012	.99
PLO0AVO	-124.18	66.406	-.153	-1.87	.064
DEMPC0AVO	80.283	74.574	0091	1.077	.284
AVO1AVO	**-521.93**	**203.098**	**-.527**	**-2.57**	**.011**
AVO1DEMPC	-75.135	171.496	-.055	-.438	.662
DEMPC1AVO	41.348	122.703	.062	.337	.737
DEMPC1DEMPC	**-387.51**	**178.777**	**-.333**	**-2.168**	**.032**
PLO1AVO	221.612	376.9	.349	.588	.558
PLO1DEMPC	228.966	363.385	.254	.63	.53
PLO1PLO	147.318	410.499	.133	.359	.72
AVO2AVO	201.061	240.916	.214	.835	.406
AVO2DEMPC	74.114	233.103	.052	.318	.751
DEMPC2AVO	-239.96	870.556	-.293	-.276	.783
DEMPC2DEMPC	-307.89	851.819	-.28	-.361	.718
DEMPC2PLO	-303.98	846.36	-.297	-.359	.72
PLO2AVO	132.727	149.279	.166	.889	.376
PLO2DEMPC	75.22	142.344	.082	.528	.598
%DET	-4.729	3.245	-.141	-1.457	.148
Vector Length	10.269	18.51	.04	.555	.58
K-Means	**63.223**	**13.819**	**.401**	**4.575**	**< .001****

Note. This model tests if the predictors in the discriminant function predict team performance. Significant predictors are in bold, with asterisks denoting the following *$p <$ .01, **$p <$ .001.

3.6.2.2 TPE

The results of the regression indicated that the model explained 45.9% of the variance and that the model was a significant predictor of TPE, $F(20,129) = 7.32$, $p < .001$. The coefficients are summarized in Table 24. Based on the results, the lag 1 transition probabilities with the sequence AVO → DEMPC (B = 643.394, β = .295, t = 2.58, p =.011), DEMPC → AVO (B = 361.689β = .34, t = 2.02, p =.045) and K-Means of the cosine similarities (B = 109.317, β = .432, t = 5.43, $p < .001$) contributed significantly to the model. The final predictive model was:

Average TPE score = 604.924 + (244.506*AVO0PLO) + (101.814*PLO0AVO) + (-47.429*DEMPC0AVO) + (-171.51*AVO1AVO) + (643.394* AVO1DEMPC) + (361.689* DEMPC1AVO) + (278.783* DEMPC1DEMPC) + (-400.62* PLO1AVO) + (--327.89* PLO1DEMPC) + (-996.46*PLO1PLO) + (-550.76* AVO2AVO) + (-541.26* AVO2DEMPC) + (-1087.8* DEMPC2AVO) + (-1490.5* DEMPC2DEMPC) + (-773.62* DEMPC2PLO) + (-57.947* PLO2AVO) + (-156.8* PLO2DEMPC) + (1.075*%DET) + (2.58* Vector Length) + (109.317* K-Means)

Table 24 – Summary of Multiple Regression Analysis for TPE.

	B	*SE B*	*β*	*t*	*p*
(Constant)	1736.67	1379.3		1.259	.21
AVO0PLO	244.506	136.709	.146	1.789	.076
PLO0AVO	101.814	96.756	.078	1.052	.295
DEMPC0AVO	-47.429	108.658	-.033	-.436	.663
AVO1AVO	-171.51	295.922	-.108	-.58	.563
AVO1DEMPC	**643.394**	**249.877**	**.295**	**2.575**	**.011**
DEMPC1AVO	**361.689**	**178.784**	**.34**	**2.023**	**.045**
DEMPC1DEMPC	278.783	260.485	.149	1.07	.287
PLO1AVO	-400.62	549.16	-.393	-.73	.467
PLO1DEMPC	-327.89	529.466	-.227	-.619	.537
PLO1PLO	-996.46	598.113	-.558	-1.666	.098
AVO2AVO	-550.76	351.024	-.366	-1.569	.119

Table 24 continued.

AVO2DEMPC	-541.26	339.641	-.235	-1.594	.113
DEMPC2AVO	-1087.8	1268.44	-.828	-.858	.393
DEMPC2DEMPC	-1490.5	1241.14	-.843	-1.201	.232
DEMPC2PLO	-773.62	1233.18	-.471	-.627	.532
PLO2AVO	-57.947	217.505	-.045	-.266	.79
PLO2DEMPC	-156.8	207.401	-.106	-.756	.451
%DET	1.075	4.728	.02	.227	.82
Vector Length	2.58	26.97	.006	.096	.924
K-Means	**109.317**	**20.135**	**.432**	**5.429**	**< .001**

Note. This model tests if the predictors in the discriminant function predict team performance. Significant predictors are in bold, with asterisks denoting the following *$p <$ .01, **$p <$ .001.

Altogether, the two multiple regression models predicting team performance and TPE supported H6, indicating that several of the communication flow and content features selected from the discriminant function predicted team outcome variables: team performance and TPE.

CHAPTER 4. DISCUSSION

This study illustrates whether communication flow and content features can discriminate between human-AI teams and all-human teams and which of these features predict team performance. The present study will discuss the interpretations and implications of the findings within the context of the feature types (flow and content).

4.1 Communication Flow

The first four hypotheses were about aspects of communication flow serving as discriminant predictors for differentiating all-human teams and human-AI teams. The flow features selected were communication determinism as well as lag 0, lag 1, and lag 2 transition probabilities. In support of H1, this study found %DET and a handful of transition probabilities contributed to the first discriminant function, successfully classifying 88.0% of the team types. Unexpectedly, %DET, although being a predictor, had a minimal contribution to the discriminant function. It also failed to support H2, in which the current study predicted human-AI teams would have higher communication determinism than all-human teams. It was surprising to find that expert teams had the highest average %DET score, and the current study only found a significant difference comparing expert teams and all-novice teams. There may be a possible data loss due to choosing a discrete method to assess teams' communication determinism for getting these results.

One potential reason for the null results might be because the current study chose to compute %DET over whole 40-minute missions, rather than using a continuous windowing

procedure (e.g., Gorman et al., 2020). The current study picked the discrete approach due to previous research using this approach to distinguish between more rigid and more flexible team communication patterns (Gorman et al., 2012). The difference between the current study and the previous study was that the current study only contained intact teams, whereas the previous study included both intact and newly mixed teams. The teams in the previous study also went through more missions than the teams in the current study. Thus, the %DET computed for the current study might not capture all the differences between human-AI and all-human teams' communication flow.

A significant %DET effect found that there were communication determinism differences among the three conditions. Although it was due to the difference between all-novice and expert teams, it nevertheless provides insights into the all-human teams. The all-novice teams had the lowest average %DET, suggesting that teams that did not have someone leading the communication were more variable in their flow patterns. The expert teams had the highest average %DET; however, the teams should not be categorized as fixed or rigid. A relatively higher %DET might have indicated that expert team did not communicate randomly, because the expert had expertise in performing the pilot role and could lead the team to communicate based on a certain sequence. This does not rule out the possibility that expert teams were being flexible, because expert teams' %DET has not traditionally been considered too high to be rigid, previous research refer to this type of teams as being metastable (Demir et al., 2018).

The third hypothesis predicted that less communication would occur in human-AI teams compared to all-human teams. The current study found significant differences in communication frequency between human-AI teams and all-human teams, as well as

among the three conditions. The multiple comparison results found that expert teams communicated significantly more than both all-novice teams and human-AI teams. In addition, human-AI teams had the lowest communication frequency. Overall, the results supported H3, indicating that human-AI teams did communicate less. This may be due to practical reasons such as the AI agent's slow reading speed, or the expert teams went through more targets. The AI agent used in the current study did have a slow reading speed and could not comprehend messages that contained misspelling or incomplete sentences (e.g., missing "a" or "the"). Because of the AI agent's limited communication ability, messages took longer to transmit.

Additionally, because the photographer has to negotiate with the pilot for changing altitude and speed, slow message transmission would lead to a longer time taken at each target, thus resulting in overall lower communication frequency compared to all-human teams. Another interpretation may be plausible, given that the expert pilot could read the message at a faster speed and be capable of speeding up message transmission. Not only do team members with an expert communicate more at each target, but also communicate more because they reached more targets, resulting in more communication compared to all-novice and human-AI teams.

The fourth hypothesis predicted that human-AI teams would have longer chains compared to all-human teams. Results did not support H4 with more lag one and two chains identified in all-human teams, and max chain length did not differ between human-AI and all-human teams. A possible explanation for the null results might be because the team task for the current study was structured in a way that required interactions had to be made among team members, yet certain information was shared between at least two team

members, resulting in a flexible way to acquire information. As shown in the results, the lag one and lag two chains that either human-AI teams or all-human teams performed were what this study proposed as good communication sequences between the AI agent and the human teammates. An interesting discovery was that the photographer did not appear to have stable communication sequences in the human-AI teams, indicating that the photographer might sometimes choose to not communicate with teammates, or interact with the navigator to acquire information that originally should come from the pilot.

Together, for communication flow, the results might further suggest that human teammates may not enjoy communicating with the AI agent compared to communicating with other humans. With median %DET scores, less communication happening, and no significant lag one or lag two chains involving the photographer in human-AI teams, it is possible that the human teammates were exhausted by the limited language capability of the AI agent. Unless it was necessary, the human teammates might not reach out to the AI agent, either being passive toward the task by not communicating to the AI pilot or try to acquire information elsewhere by interacting with a human teammate.

4.2 Communication Content

The fifth hypothesis regarded which aspects of communication content might serve as discriminant predictors for differentiating all-human and human-AI teams. Results generally supported H5 with K-Means clustering based on cosine similarities (i.e., semantic relatedness between transcripts) successfully predicting 98% of the human-AI teams and 88% of the expert teams. All-novice teams were the hardest to classify because they did not cluster as well based on the semantic relatedness (cosine similarity) compared

to the two other types of teams. Nonetheless, a combination of vector length and K-Means group membership being included as predictors in discriminant function 2 increased the classification results to 91.3% correctness for classifying team type.

Results using vector length alone presented a unique picture into the communication content differences, in that expert teams engaged in more domain-specific content utterances than all-novice teams and human-AI teams. The results suggested that expert teams might be more focused on the task compared to all-novice teams and human-AI teams. Additionally, because vector length and message length are known to be highly correlated, that expert teams communicated more frequently could account for higher vector length. What all-novice teams and human-AI teams may experience was that they were not able to take good photos of targets and therefore generated communication on the failure rather than the task. Similarly, because all-novice teams and human-AI teams communicated less, this could account for lower vector length in these two types of teams.

The K-Means clustering results showed that expert teams' communication was more semantically similar to each other as were the human-AI teams. However, all-novice did not cluster well in terms of semantic similarity. The expert team cluster might suggest specific communication content that the expert might initiate (e.g., consistent patterns of pushing and pulling, positive, closed-loop communication) that differentiate these teams from all-novice or human-AI teams. Overall, this finding demonstrated that the content metrics produced by LSA were able to discriminate between all-human and human-AI teams.

4.3 Team Performance

The final hypothesis examined whether communication flow and content features can predict team performance. The present study addressed this hypothesis with regard to two performance variables, mission team performance, and target processing efficiency (TPE). For the correlation analysis, the current study only found one feature to be moderately correlated with both team performance and TPE, and that was the K-Means cluster membership assignment based on cosine similarities (semantic relatedness). For the multiple regression analysis, the current study found the model to be significant in predicting team performance and TPE. These results indicate that communication content and flow differences among conditions may also predict team performance and TPE differences across these conditions.

4.4 Limitations and Future Directions

The primary concern of the current study was whether communication flow and content features can discriminate between human-AI and all-human teams. Because the data were not collected specifically for the current study, there were only 50 human-AI teams' missions compared to 100 all-human teams' mission, minimizing the possible differences this study may find between these two general types of teams. Future research will be necessary to address these issues, by collecting the same number of teams for all-human and human-AI conditions.

Another limitation is the AI agent's "age". In other words, the data were collected several years ago with an earlier implementation of the synthetic teammate. As technology advances, it remains uncertain if improvements in the AI agent's ability, such as reading speed and communication comprehension, would affect communication flow and/or

content. Additionally, the current study only included an AI agent to fulfill the pilot role, and the impact of the AI agent fulfilling the photographer role or the navigator role is untested. Future research should look at different AI agents with heterogeneous roles to determine other communication features that are able to discriminate between human-AI teams and all-human teams to make the predictors generalizable to other types of teams.

Furthermore, the current study used existing methods to assess communication flow and content features, and other features may be added using different algorithms. For example, in creating the LSA space, the Truncated SVD function provided by scikit-learn on Python was used to decompose the document term matrix, but SVD can be performed differently, such as using the randomized SVD function of scikit-learn. Additionally, there are different ways to train K-Means clustering algorithms to achieve better classification results that could be tested. Moreover, it should be possible to build a machine learning model to learn from previous data from all-human teams and human-AI teams to classify novel team communication data as coming from a certain type of team. Therefore, future research should look into the possibility of using different algorithms or creating machine learning models to discriminate between human-AI and all-human teams.

4.5 Conclusion

The current findings illustrate that communication flow and content can be useful features for discriminating between human-AI and all-human teams. It is the current study's hope that this work can provide helpful feedback showing differences in communication flow and content between human-AI and all-human teams to provide suggestions on what communication aspects of the AI should be the focus of team

performance improvement. Furthermore, these results should generalize to other teams that have equivalent requirements on the AI's ability to perform bi-directional communication. This work is beneficial as communication is critical for teamwork. To create more effective human-AI teams, programmers need specific information on what level of language ability, including both content and flow, AI teammates should have. One practical implication of this research is to help design AI teammate communication abilities using predictive communication features from high performing (e.g., expert) teams discovered in this research.

REFERENCES

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

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge university press.

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

Brown, M. T., & Wicker, L. R. (2000). Discriminant analysis. In *Handbook of applied multivariate statistics and mathematical modeling* (pp. 209-235). Academic Press.

Chen, J. Y. (2018). Human-autonomy teaming in military settings. *Theoretical issues in ergonomics science*, *19*(3), 255-258.

Cooke, N. J., & Gorman, J. C. (2012). The pragmatics of communication-based methods for measuring macrocognition. *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams,* 162-178.

Cooke, N., Gorman, J., Duran, J., & Taylor, A. (2007). Team Cognition in Experienced Command-and-Control Teams. *Journal of Experimental Psychology: Applied, 13*(3), 146-157.

Cooke, N., Kiekel, P., Salas, E., Bowers, C., Stout, R., & Cannon-Bowers, J. (2003). Measuring Team Knowledge: A Window to the Cognitive Underpinnings of Team Performance. *Group Dynamics: Theory, Research, and Practice, 7*(3), 179-199.

Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2018). Team coordination and effectiveness in human-autonomy teaming. *IEEE Transactions on Human-Machine Systems, 49*(2), 150-159.

Dong, A. (2005). The latent semantic approach to studying design team communication. *Design Studies*, *26*(5), 445-461.

Foushee, H. C., & Manos, K. (1981). Information transfer within the cockpit: Problems in intracockpit communications. In C. E. Billings & E. S. Cheaney (Eds.), Information transfer problems in the aviation system *(Report No. NASA TP-1875)*. Moffett Field, CA: NASA-Ames Research Center.

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

Gorman, J., & Cooke, N. (2011). Changes in Team Cognition After a Retention Interval: The Benefits of Mixing It Up. *Journal of Experimental Psychology: Applied, 17*(4), 303-319.

Gorman, J., Cooke, N., Amazeen, P., Fouse, S., Duchon, A., Keyton, J., & Miller, A. (2012). Measuring Patterns in Team Interaction Sequences Using a Discrete Recurrence Approach. *Human Factors: The Journal of Human Factors and Ergonomics Society, 54*(4), 503-517.

Gorman, J. C., Grimm, D. A., & Dunbar, T. A. (2018). Defining and measuring team effectiveness in dynamic environments and implications for team ITS. In *Building Intelligent Tutoring Systems for Teams*. Emerald Publishing Limited.

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

Homan, A. C., Van Knippenberg, D., Van Kleef, G. A., & De Dreu, C. K. (2007). Bridging faultlines by valuing diversity: diversity beliefs, information elaboration, and performance in diverse work groups. *Journal of applied psychology*, *92*(5), 1189.

How, J. P. (2016). Human-Autonomy Teaming [From the Editor]. *IEEE Control Systems Magazine*, *36*(2), 3-4.

Jentsch, F. G., Salas, E., Sellin-Wolters, S., & Bowers, C. A. (1995, October). Crew coordination behaviors as predictors of problem detection and decision making times. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 39, No. 20, pp. 1350-1353). Sage CA: Los Angeles, CA: SAGE Publications.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, *104*(2), 211.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes, 25*, 259–284.

Marathe, A. R., Schaefer, K. E., Evans, A. W., & Metcalfe, J. S. (2018). Bidirectional communication for effective human-agent teaming. *Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10909*, 338-350.

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of management review*, *26*(3), 356-376.

Marlow, S. L., Lacerenza, C. N., Paoletti, J., Burke, C. S., & Salas, E. (2018). Does team communication represent a one-size-fits-all approach?: A meta-analysis of team communication and performance. *Organizational Behavior and Human Decision Processes*, *144*, 145-170.

Mcneese, N., Demir, M., Cooke, N., & Myers, C. (2018). Teaming With a Synthetic Teammate: Insights into Human-Autonomy Teaming. *Human Factors: The Journal of Human Factors and Ergonomics Society, 60*(2), 262-273.

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of applied psychology*, *94*(2), 535.

Mosier, K. L., & Chidester, T. R. (1991). Situation assessment and situation awareness in a team setting. *Designing for everyone*, 798-800.

O'Connor, B. P. (1999). Simple and flexible SAS and SPSS programs for analyzing lag-sequential categorical data. *Behavior Research Methods, Instruments, & Computers*, *31*(4), 718-726.

Orasanu, J. (1990). *Shared mental models and crew performance* (Report No. CSLTR-46). Princeton, NJ: Princeton University.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, *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). Westport, CT, US: Ablex Publishing.

Scalise, R., Bisk, Y., Forbes, M., Yi, D., Choi, Y., & Srinivasa, S. (2018). Balancing Shared Autonomy with Human-Robot Communication.

Shively R.J., Lachter J., Brandt S.L., Matessa M., Battiste V., Johnson W.W. (2018) Why Human-Autonomy Teaming?. In: Baldwin C. (eds) Advances in Neuroergonomics and Cognitive Engineering. AHFE 2017. Advances in Intelligent Systems and Computing, vol 586. Springer, Cham

Urban, J., Bowers, C., Monday, S., & Morgan Jr, B. (1995). Workload, Team Structure, and Communication in Team Performance. *Military Psychology, 7*(2), 123-139.

Webber, C. L., Jr., & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology, 76*, 965–973.

Wynne, K. T., & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, *19*(3), 353-374.