The Effect of Individual Traits on Emerging Roles in Synchronous Computer-Mediated Groups

VISHAL B. VATNANI, California State University, Fullerton, USA
JORDAN B. BARLOW, University of St. Thomas, USA

This study examines group dynamics in a virtual decision-making context through the lens of role theory and speech act theory, furthering previous research on the roles that emerge in such settings. Specifically, this research uses communication transcripts and text analysis to complete two main research objectives. First, we extend prior research by exploring how individual traits of group members (specifically, personality, cognitive ability, and social sensitivity) are related to emerging roles in computer-mediated communication. Personal characteristics such as agreeableness, social sensitivity, and motivation were significant predictors of role emergence in online decision-making groups. Agreeableness predicted the emergence of the listener role, social sensitivity predicted the emergence of the manager role, and motivation predicted the emergence of the sharer role. Second, we answer the call for more replication research and further validate the results from previous research by measuring roles assumed by individuals working synchronously online; we highlight the findings that are confirmed across previous research and our current study.

CCS Concepts: • **Human-centered computing → Collaborative and social computing**

Additional Key Words and Phrases: Computer-mediated communication, roles, speech act theory, personality, virtual work

ACM Reference format:
Vishal B. Vatnani and Jordan B. Barlow. 2024. The Effect of Individual Traits on Emerging Roles in Synchronous Computer-Mediated Groups. *Proc. ACM Hum.-Comput. Interact.*, 8, CSCW1, Article 181 (April 2024), 22 pages, https://doi.org/10.1145/3641020

1 INTRODUCTION

In social interactions, individuals assume one or more roles [1]. Roles are defined as a set or pattern of behaviors characteristic of a person in a given context [2]; such behaviors may rise from expectations of oneself or others [3]. Role theory is based on the principle that patterns of behavior exhibited by an individual are largely based on two aspects [2]: an individual's social identity or position, and the context or situation in which the individual finds themselves. For example, an individual may carry out the role of a "student" (identity) in a classroom (context), given their enrollment for a course. Previous research has examined employee roles by studying patterns of behavior in the workplace and their effects on employee outcomes. For example,

Author's addresses: V. Vatnani, California State University, Fullerton, 800 N. State College Blvd. Fullerton, CA 92831, USA; J. Barlow, University of St. Thomas, 2115 Summit Ave., St. Paul, MN, USA

2573-0142/2024/4 – 181

https://doi.org/10.1145/3641020

Welbourne, Johnson and Erez [1] study employee performance based on role theory and identity theory, and define multiple roles such as "Career" and "Innovator."

In group decision-making, particularly in self-managing groups, informal communication roles may emerge if no formal roles are defined. Group decision-making is particularly interesting to researchers and practitioners in information systems and computer-mediated communication (CMC) [4, 5]. Such decision-making groups are increasingly virtual, especially during and after the COVID-19 pandemic [6]. Much organizational communication happens via text tools including e-mail, text messaging, and instant messaging. Even when groups use video or audio to meet via collaboration and communication technologies such as Teams, Zoom, and Slack, individuals will often supplement these meetings with built-in conversation threading, instant messaging, and within-call chat capabilities of these tools [7]. While past research on face-to-face groups gives some insight into emergent roles [e.g., 8], virtual decision-making groups have characteristics distinct from offline settings, including language use, synchronicity, speed of response, parallelism, etc. [9].

Several studies in the CMC literature examine individuals' roles by measuring interaction and speech patterns. Earlier research on CMC role emergence has examined roles mainly in online communities and other virtual asynchronous settings [e.g., 10, 11, 12]. Very few studies [e.g., 13, 14, 15] have focused on the emergence of multiple roles in synchronous online communication.

In the present research, we examine emergent roles and their relationship with individual traits in a text-based CMC environment. Our previous work [14] suggests that four granular roles (*managers*, *sharers*, *listeners*, and *opinionators*) dynamically emerge in such decision-making tasks. This suggests that, in a virtual setting where roles are not formally defined, roles are more fluid. The present study takes the perspective of roles being emergent rather than being explicitly assigned. We extend previous research on emerging roles in synchronous CMC by pursuing the following research questions:

RQ1. Do individual traits predict the emergence of roles in text-based synchronous CMC?
RQ2. Can previous roles in synchronous CMC be further replicated and validated?

To answer the first RQ, our main research objective is to understand how individual characteristics are related to the emergence of roles in group-based decision-making CMC tasks. Role theory suggests that the pattern of individual behavior (i.e., role) a person assumes is a function of context, interaction, and individual attributes [16]. In other words, we can measure roles by examining behavior, but to truly understand those roles we must understand both the social interaction and individual attributes that affect such roles. Specifically, this study examines the effects of three types of individual characteristics (personality, cognitive ability, and social sensitivity) on the emergence of communication roles in synchronous CMC.

To answer the second RQ, our additional research objective is to replicate the findings of Barlow [14, 15] to further validate this work in support of the first research question. Following the procedures of Barlow [14, 15], we encode written text to speech acts and then cluster such speech acts, along with message characteristics such as frequency and length, into roles. In other words, by studying the meaning and context of each message sent in a group chat given a decision-making task at hand, we group these meanings into emergent roles. Scholars in communications [17], information systems [18], and many other fields are calling for additional replication research to improve confidence in published research studies. Specific to the literature on roles in CMC, many studies explore and define roles in various contexts, synchronous and asynchronous, but such roles should be further validated, particularly when extending research on those roles. In other words, research findings should be thoroughly validated before researchers attempt to extend them. Because we believe it is important to understand how individual traits can predict roles, and we study this research question in the context of decision-making groups using CMC, we also validate the roles that have been proposed in previous research in this context. Thus, 3this paper presents both replication and extension research. This is not to discount the value, and indeed the need, to explore additional task types, contexts, and communication environments; we believe it is necessary to both explore additional contexts as well as to validate the definition of roles previously set forth in prior research.

2 THEORETICAL FRAMEWORK

2.1 Role Theory and Speech Act Theory

Humans behave in predictable ways depending on their social identities and context [8]. Role theory defines roles as characteristic behavior patterns, where people hold expectations for their own and others' behavior [8]. Thus, the major generator of roles according to most role theorists is expectations. Such expectations may arise due to factors such as personal characteristics, environment, social positions, function of roles that other people assume, and socialization [2]. This theory also informs multiple research studies on emergent roles in computer-mediated settings [e.g., 11, 19, 20].

In online groups and communities, roles are often not defined, yet roles emerge—individuals are likely to respond to group or community needs and adopt the role that seems most appropriate [21]. For example, a study examining emergent roles in Wikipedia used a framework of emergent dynamics instead of a framework of organizational structure due to the constant need for change in such online communities [22]. Others have found that individual roles emerge in virtual teams when role ambiguity is high and task interdependency is low [23]. Similarly, this research examines how roles emerge in a CMC setting with a decision-making task, even when no formal roles are defined.

In accordance with role theory, communication is the basis for understanding roles that emerge in CMC. The great majority of studies analyzing CMC roles have examined roles that emerge in asynchronous platforms, including Wikipedia [e.g., 24, 25], social media [26-28], and other online communities [29-31].

The roles detected by researchers vary by context. For example, in asynchronous collaborative learning environments, roles such as 'encourager,' 'dominator,' and 'fellow-traveler' emerge [32]. Roles that emerge in Wikipedia [20, 22, 33, 34] include 'substantive experts,' 'vandal fighters,' 'social networkers,' 'watchdogs,' and many others. Such roles are not appropriate for settings such as synchronous decision making [14].

Few studies have examined roles that emerge in synchronous online communication. One exception is Dowell, Nixon and Graesser [13], who examined online group conversations in collaborative learning environments. They found the following roles emerging: 'driver,' 'chatterer,' 'lurker,' 'follower,' 'socially detached,' and 'influential actor.' A few studies have examined communication patterns in synchronous groups [35-37], though these studies do not examine roles and role emergence.

Barlow [14, 15] was one of the first to define roles in a synchronous CMC setting, particularly for decision-making contexts. We argue that, in addition to role theory, speech act theory is another critical lens through which to understand CMC discourse, especially synchronous CMC. Speech act theory, grounded in work by Austin [38], conceptualizes a statement as an act of speech—"doing things" with words [39]. Many studies do not focus heavily on message content to understand roles; rather, they focus on how much and with whom individuals communicate in asynchronous settings [e.g., 11, 19, 24, 26, 28-30, 40-45]. However, according to speech act theory, behavior can only be truly understood by understanding the meaning of communicative acts.

Only a handful of CMC role studies examine the content or meaning of communication rather than focusing on the amount, patterns, or interaction of messages. A few studies [13, 46] use text mining to examine word patterns, similarities of messages, newness of messages, etc., but do not consider types of speech acts (i.e., the types of behaviors exhibited by participants). In addition, a few studies have used content or conversation analysis [47, 48] to understand message content, but these are context-based analyses of specific information being shared. Research is needed that examines CMC behavior by examining speech act types being used in CMC interaction.

The previous Barlow [14, 15] studies were the first to use speech act theory to understand roles in CMC settings. These studies examined the communication patterns (message frequency, message length, and speech act categories) of participants in a synchronous decision-making task. Surprisingly, these are the only prior studies, to our knowledge, on role emergence in CMC that are grounded in a theory that emphasizes the meaning of communicative acts. Using speech act analysis and cluster analysis, the studies found that four roles emerged in this setting. *Listeners* are those who share fewer amounts of messages, a significant portion of which are simply agreeing with others. Listeners provide moderate amounts of factual information and do little to guide the discussion. *Sharers* use most of their messages to share information with group members and vary in how much other communication they share with the group. *Managers* are those whose primary messages ask questions, structure the decision-making task, or manage the flow of discussion, but also contribute moderately by providing opinions and agreeing with others. Finally, *opinionaters* are those who primarily share opinions about the task or information and share fewer messages with objective information. Opinionaters tend to have the longest messages.

Beyond replicating and validating previous work on roles in synchronous CMC, our main research objective is to examine how these roles relate to individual characteristics of group members—namely, personality, cognitive ability, and social sensitivity.

2.2 Individual Traits as Antecedents of Emerging Roles

Several studies indicate the importance of understanding how individual traits relate to behavior in group and/or decision-making settings. For example, Morgeson, Reider and Campion [49] studied how the selection of individuals based on their social skills, personality characteristics, and knowledge would influence team outcomes. In the subsections below, we summarize current research of personality, cognitive ability, and social sensitivity in group decision-making settings. No studies, to our knowledge, have examined the effects of such individual traits on role emergence in CMC settings; however, if traits can predict emerging roles in team settings, this knowledge would be helpful to organizations and managers that form teams, in order to better facilitate successful team projects. Because one of our research objectives is to replicate the previous Barlow [14, 15] studies, we did not assume the roles that would emerge *a priori*; thus, we did not develop specific hypotheses as to which traits would correlate with which roles. We expect some significant correlations between roles and traits, but our data analysis examines (1) which roles emerge and then (2) how those roles (and speech acts) relate to traits.

2.2.1. Personality. The Five Factor model is an organization of personality traits into five dimensions: *extraversion*, *agreeableness*, *openness to experience*, *conscientiousness*, and *neuroticism* [50]. Several researchers have examined the impact of these five personality traits on various individual and team-level outcomes. For a comprehensive background on personality research in face-to-face teams, we refer the reader to Bell [51], Curşeu, Ilies, Vîrgă, Maricuţoiu and Sava [52], Peeters, Van Tuijl, Rutte and Reymen [53]. A few studies examine the impact of personality on virtual team performance [54-56]. These studies have found, contrary to much research in the face-to-face literature, that extraversion does not lead to higher team performance, while one study also finds that both the mean and variance of conscientiousness among virtual team members impacts team performance [55].

Outside of team performance, little research has examined other effects of personality on team interactions, with a few exceptions. Balthazard et al. [54, 57] found that extraversion leads to constructive interaction styles, characterized by concern for both personal and group outcomes, effective information exchange, cooperation, and respect, but also concluded that extraversion is not predictive of perceived transformational leadership in virtual teams, as it would be in a face-to-face setting. Straus [58] examined the relationship between individual participation in computer-mediated group discussions and extraversion. Gutiérrez Maldonado, Mora, García and Edipo [59] studied the relationship between CMC styles and personality, and found that in virtual settings, introverts contribute as many messages as extroverts. Specifically, they found a relationship between introversion and metacognitive expressions and value statements. Taken together, these studies find that personality matters in computer-mediated communication, but there remains a lack of comprehensive research in this area.

We hypothesize that personality will be related to the roles that emerge in computer-mediated decision-making groups, such that one or more personality traits are significantly correlated with one or more speech acts/roles.

2.2.2. Cognitive Ability. Cognitive ability is the ability of an individual to perform well across multiple types of tasks and is sometimes called general mental ability or intelligence [60]. Past research in face-to-face teams has indicated that individual cognitive ability does influence the performance of teams [51, 61]. However, this effect tends to be small and is often understood better when studied in conjunction with other traits such as personality [56, 62].

Little research has studied cognitive ability in CMC teams [63], and this research indicates that individual cognitive ability helps teams perform better in brainstorming tasks [64] but not in decision-making tasks [56, 65]. These studies have focused on the overall performance of teams but have not examined how cognitive ability affects roles and interactions among group members. Research in face-to-face teams has indicated that group members' expectations and perceptions of cognitive ability interact with perceptions of personality traits [66]. Given that roles are based on expectations of self and others, we anticipate that cognitive ability will be related to the roles that emerge in computer-mediated decision-making groups.

2.2.3. Social Sensitivity. Social sensitivity is an ability to understand feelings and emotions of others [67]. Social sensitivity is often measured through the "Reading the Mind in the Eyes" test, which asks individuals to identify emotions by looking at photographs of only the facial area containing eyes and eyebrows [68, 69].

Because social sensitivity is measured with visual cues, which are not present in many CMC settings, there is debate in the literature about the its applicability in online settings; however, Engel, Woolley, Jing, Chabris and Malone [67] found that even in virtual settings with no visual cues present, social sensitivity was a key factor in determining the success of a team. Regardless of measurement, sensing the emotions of others is a key aspect in working together as a team. Like personality, we believe that such traits that typically affect group interaction should affect the roles that people assume when participating in a team. Thus, we hypothesize that social sensitivity will be related to roles that emerge in computer-mediated decision-making groups.

3 METHODS

3.1 Tasks and Procedures

The study consisted of two phases: one that involved working in randomly assigned groups of three to five members in a campus lab to complete a task, and the other where the participants completed a survey individually online. The task to be completed by the groups was a modified version of the university admissions task, which has been used in many previous studies [e.g., 70, 71]. Because we aimed to replicate the previous Barlow [14, 15] studies, we chose to use the same task to maintain the same context.

In this task, groups were instructed to make university admissions decisions regarding a set of eight possible candidates; groups were allowed to admit up to four. Participants sat in individual cubicles where they could not see physically who was in their group. They communicated using the chat function of Google Hangouts and used a Google spreadsheet provided to them as a workspace to record information and submit their answer. In other words, although students were collocated at the time of the experiment, all groups worked online and not in a face-to-face format.

Groups were given 20 minutes to complete the task; a time limit was only provided to ensure consistency between groups, and pilot testing of the task showed that 20 minutes is more than sufficient time to choose a solution. After the task, participants completed an online survey asking for demographic information, responses to the personality measure items, and control variables, as detailed in the next section.

3.2 Survey Measures

Personality was measured using the Big 5 Personality scale items [72]. The Big 5 scale has 44 items (8-10 items for each trait). For each item, participants were asked to respond on a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) whether they felt they matched the personality aspect described in the item. Individual scores were calculated by aggregating the responses to the 8-10 items, resulting in a score of 1-7 for each personality trait.

Cognitive ability was measured using the Wonderlic Personnel Test (WPT-Q) [73, 74]. This eight-minute assessment contains 30 questions, and the scores are converted to a 0-50 scale. WPT items are proprietary and cannot be listed in our manuscript; sample questions are available from Wonderlic's website.

Social sensitivity was measured with the Reading the Mind in the Eyes test [68]. This assessment contains 36 questions where participants must select the emotion displayed in a displayed photograph of eyes. Each question has an objectively correct answer; thus, the score ranges from 0-36.

In addition to these traits, we also collected the following as control variables: gender, age, whether the participant was a native English speaker, motivation on the task (single-item measure developed by the authors: "For the College Admissions exercise, I made an effort to complete the task effectively."), and perceived mutual trust (items adapted from Edmondson [75] and Fransen, Kirschner and Erkens [76]). We do not believe that individual traits alone predict communication patterns and roles; other factors could affect the way group members communicate and act. During the design of the study, we anticipated based on our experience working in teams that motivation and trust could potentially impact how participants communicate in a team setting; however, they are not individual traits, which was the focus of our study, so we include them only as control variables.

3.3 Participants

Participants were 191 undergraduate and 64 graduate students enrolled in various degree programs at the business school of a large U.S. public university. Participation was incentivized by extra-credit awarded to participating students upon completion of both an on-campus lab experience and an online survey. We ran 28 lab sessions with 255 students in 62 teams. 248 students successfully completed the on-campus task; 220 also completed the survey. 109 (49.5%) were female, 110 (50.0%) were male, and 1 (0.5%) declined to respond regarding gender. The average age of undergraduate students was 23.6 years (SD = 4.8) while the average age of graduate students was 25.4 years (SD = 2.5). The average age of the total population was 24.3 years (SD = 4.4). Three participants did not complete the age question, two participants did not complete the cognitive ability assessment questions, and one participant did not answer the gender question. Missing age, gender, and cognitive ability scores were imputed using the Multiple Imputation by Chained Equations method (MICE). The MICE method works by iteratively making multiple predictions for missing values in each variable using the complete data [77]. This method assumes that data are Missing At Random (MAR), which we believe is the case with the online survey, and that the data follow the assumptions made by regression models. Not treating data that are MAR can lead to biased estimates.

3.4 Computer-Mediated Discourse Analysis

One approach used for studying text-based behavior in an online environment is Computer-Mediated Discourse Analysis (CMDA). What identifies a textual analysis as CMDA is that the analysis tries to understand online behavior that is grounded in empirical, textual observation [39]. The CMDA approach can be applied to five levels of linguistic analysis: (1) structure (e.g., sentence and word formations); (2) meaning (e.g., meanings of words, speech acts); (3) interaction (e.g., turn-taking, topic development); (4) social (e.g., expressions of play, conflict, power, group membership); (5) participation patterns (e.g., message length, frequency) [39].

In studies of asynchronous CMC roles, researchers often use (1) network analysis to analyze the *interaction* level of linguistic analysis, which is concerned with communication collaboration patterns, and (2) participation analysis to analyze the *participation* level of linguistic analysis, which includes descriptive measures such as the number of messages and replies exchanged, message length, etc. [e.g., 19, 41, 42]. However, researchers have called for more studies on CMC roles to incorporate analyses at the *meaning* level of text concerned with the meaning of messages and how language is used in context [46]. Especially in a synchronous context, speech act analysis is a better fit both theoretically and methodologically. Thus, to understand behavior in mediated groups through textual analysis on the *meaning* level of CDMA linguistic analysis, we will try to understand speech acts. Like other studies of CMC roles (both synchronous and asynchronous), we also use participation analysis to analyze the *participation* level. Participation and meaning of communication are both essential in understanding roles in synchronous CMC.

Participation analysis is commonly used in studies of asynchronous CMC roles [e.g., 10, 78] and examines communication or other behaviors relevant to the CMC platform. Examples of participation analysis for platform-specific behaviors include number of lines of code edited on GitHub [41], number of gameplay actions taken in an esports platform [79], number of retweets on Twitter [43] or lending activities on a micro-lending platform [12]. Because this study focuses on a decision-making task where no other online behaviors were involved, only participation analysis of communication acts was used. The participation measures used were (1) percentage of the group's messages by each participant, and (2) length (in words) of each message.

Speech act analysis is a technique based on speech act theory [38, 80] that has been used in research in various disciplines [81]. In speech act analysis, one reads every message in a communication record and assigns it to a speech act category, which describes the type of behavior depicted in the message (e.g., "question," "reaction").

Participants in the study exchanged 6305 messages in total, and 6136 unique messages. We consider each message to be a speech act. We first coded 630 (about 10%) messages selected using simple random sampling. We used the "CMC act taxonomy" created by Herring, Das and Penumarthy [82] and validated in subsequent research [e.g., 83]. This taxonomy defines 16 types of speech acts. For consistency with the previous studies [14, 15], we combine INQUIRE, DIRECT, MANAGE, and REQUEST speech acts into the MANAGE speech act category at the time of coding, leaving 13 speech act types that could be assigned to any message in our transcripts. Eight of these were commonly used in our data; these are defined with examples in Table 1. The remaining five were used infrequently.

Each of the two authors of this paper independently coded the 630 messages into one of the eight speech acts, and the OTHER speech act for messages that couldn't be classified into one of the eight speech acts. We assessed the interrater reliability between the two authors and obtained Cohen's kappa coefficient of 0.69. The authors did not agree on 152 messages and later met and resolved the disagreements in classifications through discussion and exchanging insights to find a common ground.

Because of the large number of messages in our data set, we used text mining to predict and assign speech acts on the remaining 5506 messages. We used the LightSide text mining tool [84] to train a regression model and assign a specific speech act type to each transcript line. The 630 coded messages were used as the training set for this model. We condensed the GREET, REACT, OTHER, REPAIR, and THANK speech acts into OTHER because they made up less than 6% of the total coded speech acts and could lead to class imbalance when used in the training set as one of the target categories. The final speech acts used in the training set were ACCEPT, CLAIM, INFORM, MANAGER, and OTHER.

Table 1. Subset of CMC Speech Act Taxonomy by Herring, Das and Penumarthy [82]

Speech act	Definition (quoted from Herring, Das and Penumarthy [82])	Example from our data
INFORM	Provide "factual" information, verifiable in principle, even if untrue; inform, state	*we have to select 4 answers*
MANAGE	Manage discourse, organize, prompt, focus, open or close discussions	*so what are your answers?*
CLAIM	Make a subjective assertion, unverifiable in principle; assert, guess, speculate	*It makes little difference that they have not taken them prior to application.*
ACCEPT	Concur, agree, acquiesce	*Ok i agree with AGH too*
REACT	Show listenership, engagement (positive, negative, or neutral), endorse, approve	*and I didn't even ask a question lol*
THANK	Express gratitude, appreciation, or acknowledgement towards another individual or the team	*Thank you to whoever suggested using the spreadsheet!*
GREET	Greeting, leave taking, inquiries about/wishes for well-being	*Hi everyone*
REPAIR	Return, clarify, correct misunderstanding	*oh wait... we didnt share anything about the candidates*

The model for the prediction task employed the logistic regression algorithm with L2 regularization because it gave us the best performance when running a ten-fold cross validation on the training set over other models. We extracted the following features from the text: unigrams, bigrams, parts of speech bigrams, line length, count occurrences, punctuation, and Stem-N-Grams. We obtained an accuracy of 0.72 (kappa 0.63) for this trained model using a 10-fold cross-validation technique. The predicted 5506 messages and the initially coded 630 messages used as the training set were combined to obtain a total of 6136 messages used for further analysis.

3.5 Role Clustering

The distribution of the total speech act labels in our dataset is shown in Figure 1 below.

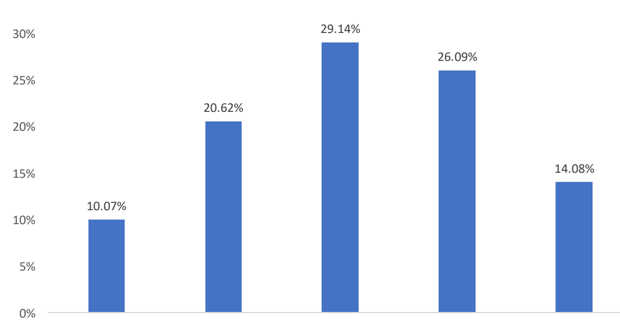

Fig. 1. Relative frequencies of speech act types

The statistical software R version 4.0.3 (2020-10-10) was used for all data analysis and visualizations relating to role clustering and regression analysis.

To aggregate the data on a participant level, we aggregated the total number of predicted speech act labels for each line of message for a participant and calculated the proportion of each speech-act to the sum of speech acts for the participant.

In our analysis, we wanted to find groups of participants who shared common characteristics across the percentage of each speech act employed (*ACCEPT*, *CLAIM*, *INFORM*, and *MANAGE*), along with the percentage of messages contributed (*Message Contribution*) to the team discussion and the average length of messages sent (*Average Message Length*). We chose to not include the percentage of OTHER speech act as a feature in further analysis even though it accounted for 14.08% of total speech acts because its interpretation is not meaningful for the purposes of this study. Together, these six features were used as an input to cluster analysis. The features used for clustering are shown in Table 2 below.

Table 2. Features Used in Cluster Analysis

Feature	Description
Message Contribution	Messages sent by an individual divided by total number of messages sent by all individuals in the team (i.e., percentage of messages)
Avg Message Length	Total words used by an individual divided by number of messages sent
% Accept	Messages predicted to be ACCEPT divided by total messages sent
% Claim	Messages predicted to be CLAIM divided by total messages sent
% Manage	Messages predicted to be MANAGE divided by total messages sent
% Inform	Messages predicted to be INFORM divided by total messages sent

Of the 248 subjects who participated in the task, we were left with 215 subjects after performing outlier removal using the Interquartile Range (IQR) and performing listwise deletion on rows with at least one of the six features being an outlier.

We normalized each feature above using min-max normalization except for Message Contribution, as each subject's contribution to the team adds up to 100%. In addition, we decided to place lesser importance on the normalized Average Message Length feature for clustering and assigned a weight of 0.5 to it by multiplying the feature by $\sqrt{0.5}$. We placed less importance on this feature for clustering because theoretically, the length of the message is not as important as its meaning. In addition, in our initial regression analysis of cluster membership as the dependent variable against the normalized independent variables used for clustering, the average message length had one of the highest coefficients for each of the four clusters (roles). This suggested that we were primarily separating the clusters based on average message length, and we wanted to reduce the weight of this feature in the clustering.

4 ANALYSIS

4.1 Role Clustering

We ran the k-means algorithm [85] to validate the number of clusters using the elbow method [86], and found a partially visible "elbow" in our plot at k=4, after which the Total Within-Cluster Sum of Squares have a smoother drop in values, as shown below in Figure 2.

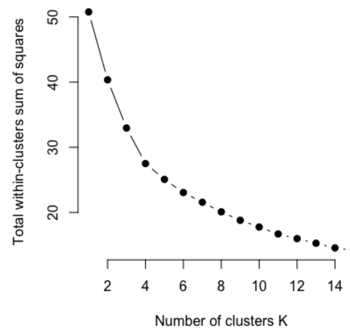

Fig. 2. Elbow plot

Through Ward's method of hierarchical clustering using complete linkage, we visually found the emergence of at least four significant clusters that provided the appropriate trade-off between the within-cluster sum of square distance of these clusters and the total number of clusters. Further, we cross-validated the number of clusters obtained by running silhouette analysis [86], which measures the distance of a cluster member with neighboring clusters to find the optimal number of clusters. The results of silhouette analysis are shown in Figure 3(a).

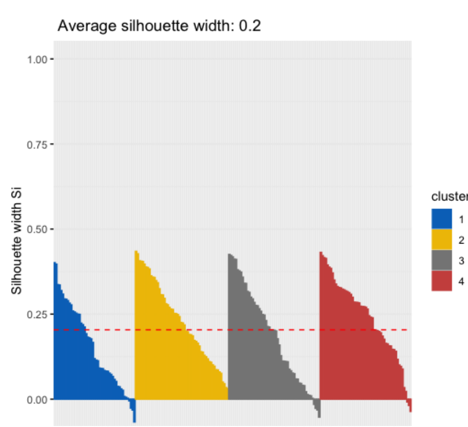

Fig. 3(a). Silhouette plot for c-means with four clusters

It is easy to interpret that the average silhouette width of 0.2 is not very large, indicating that the observations are not very well clustered and may hold multi-cluster membership. In other words, many participants assumed multiple roles. However, when the silhouette width of each cluster was compared in the case of three clusters versus four clusters, or even five clusters, we found that the size of each cluster and its silhouette width were more evenly distributed in the four-cluster solution.

The silhouette plots with three and five clusters are shown in Figure 3(b) and 3(c) below. The average silhouette width of the three-cluster solution is 0.19 with cluster-1 width = 0.16, cluster-2 width = 0.16, and cluster-3 width = 0.26 (SD = 0.047). The average silhouette width for the four-cluster solution is 0.2 with cluster-1 width = 0.15, cluster-2 width = 0.23, cluster-3 width = 0.19, and cluster-4 width = 0.24 (SD = 0.035). The average silhouette width for the five-cluster solution is 0.15. The lesser standard deviation and higher mean width with four clusters helped us decide that the four-cluster solution was the best solution.

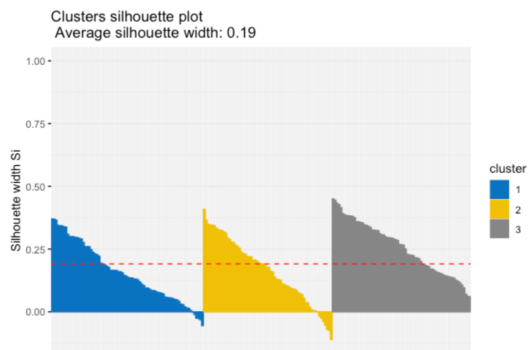

Fig. 3(b). Silhouette plot for c-means with three clusters

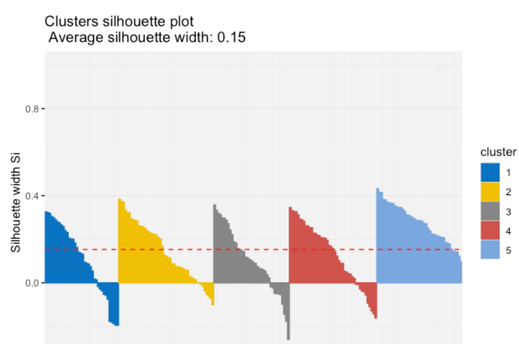

Fig. 3(c). Silhouette plot for c-means with five clusters

Given that the task was the first time that participants worked in these team configurations, it is likely roles emerged during interaction. That is, few participants would assume a role at the beginning of the exercise. It would be possible for participants to either strongly adhere to a single role or behave in a way that fits multiple roles. Previous research has not examined the strength of given roles; that is, individuals may strongly assume a single role, or they may act according to a variety of roles. We examine the strength to which an individual fits a given role. For each participant, we calculated the strength of each role; specifically, to retain partial cluster membership, we decided to use the Fuzzy c-means clustering algorithm [87] and assign non-negative partial cluster membership to each observation across the four clusters such that each cluster membership sums to 1.

Table 3 shows summary statistics for the clusters resulting from the cluster analysis. Distinguishing values are highlighted in light (low) and dark (high) gray. Figure 4 shows the Fuzzy c-means cluster plot along the first two principal components, explaining 49.2% of total variance.

4.2 Traits and Roles

The personality, social sensitivity, and cognitive ability assessments were administered online, separately from physical attendance at the lab study. Thus, not all participants were able to complete both assessments.

Table 3. C-means Cluster Analysis – Averages for Each Cluster

Cluster	Contri-bution	Avg Length	% Accept	% Claim	% Manage	% Inform	Size
Opinionater	24%	6.6	11%	36%	22%	19%	49
Manager	27%	7.0	5%	19%	41%	23%	56
Listener	22%	5.5	19%	14%	22%	29%	55
Sharer	26%	6.9	6%	18%	19%	42%	55
Total	25%	6.5	10%	22%	26%	28%	215

Note: Highest values in each column are dark gray. Lowest values in each column are light gray.

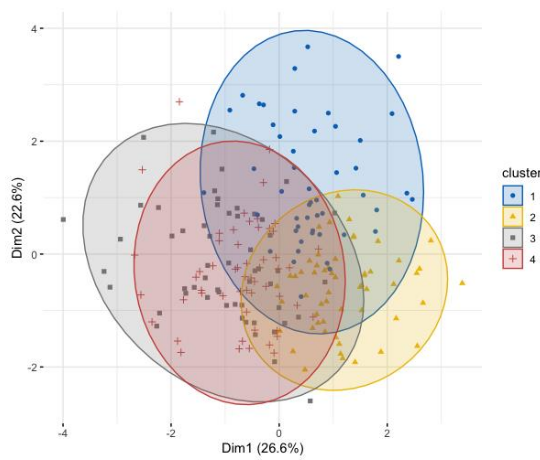

Fig. 4. Fuzzy cluster plot for c-means

By using listwise deletion, we were left with 220 participants completing both assessments from the original 248 participants who had participated in the lab study. Of the 220, six participants were missing values in their survey responses. We imputed missing data in the profile variables using the MICE method to preserve 220 rows. We also ran the analysis with listwise deletion (n=214) as an alternative and found the results to be similar. Joining the profile-variables data with the cluster-analysis data (n=215), we preserve the 190 rows for further analysis.

We calculated Cronbach's alpha for all multi-item survey constructs to establish construct reliability and found all values to be acceptable. Specifically, we calculated an alpha score for extraversion (0.861), agreeableness (0.782), conscientiousness (0.795), neuroticism (0.805), openness to experience (0.747), and mutual trust (0.799). All other items collected via survey were either single-item questions (e.g., age, gender, motivation) or items calculated based on the sum of correct answers (e.g., cognitive ability, social sensitivity), where a traditional reliability analysis would not be applicable.

We performed a multiple regression analysis using R to analyze which individual traits and characteristics were significantly predictive of the roles that participants assumed during the task. The dependent variable was the strength to which a participant corresponded to a specific cluster. During c-means cluster analysis, for each participant we calculated a number corresponding to how closely they fit the cluster/role (one number per role, summing to 1). The independent variables included personality traits and control variables. Results are shown in Table 4.

Table 4. Regression Analyses with Roles as Dependent Variables

Dep Variable:	Sharer		Opinionater		Manager		Listener	
Ind Variable	β	p	β	p	β	p	β	p
Intercept	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Personality								
Agreeableness	-0.131	0.147	0.056	0.554	-0.144	0.115	0.216*	0.019
Conscientiousness	-0.152	0.094	0.152	0.107	0.018	0.847	0.002	0.984
Extraversion	0.064	0.427	-0.023	0.783	0.104	0.202	-0.141	0.086
Neuroticism	-0.096	0.253	0.043	0.620	0.055	0.521	0.007	0.935
Openness	-0.011	0.895	0.054	0.517	-0.102	0.211	0.058	0.475
Cognitive Ability	0.067	0.377	0.045	0.570	-0.008	0.916	-0.099	0.203
Social Sensitivity	-0.030	0.711	-0.121	0.148	0.201*	0.014	-0.053	0.513
Age	0.202*	0.012	-0.022	0.789	-0.036	0.656	-0.148	0.067
Gender (f)	-0.122	0.114	0.042	0.600	0.111	0.153	-0.019	0.804
Native English	-0.104	0.168	0.014	0.859	0.041	0.593	0.053	0.485
Motivation	0.185*	0.043	-0.076	0.419	0.038	0.677	-0.153	0.098
Mutual Trust	-0.123	0.159	-0.002	0.981	-0.046	0.607	0.167	0.061
Adjusted R-squared	0.056		-0.024		0.030		0.027	
Multiple R-squared	0.116		0.048		0.092		0.088	

*p<0.05; **p<0.01

Given that we ran multiple models, we report 95% confidence intervals of the regression coefficients by cluster as shown in Figure 5.

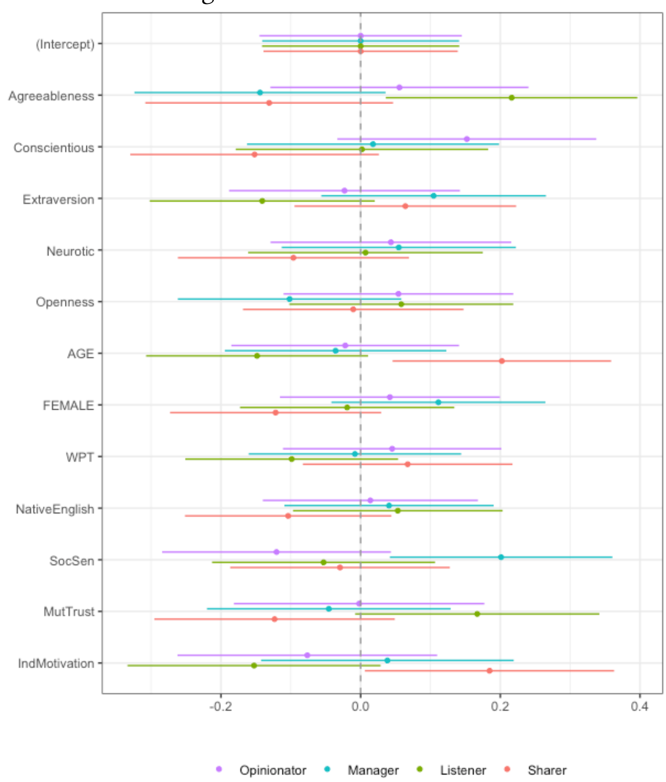

Figure 5. 95% Confidence Intervals for each independent variable

The R function call used for the regression analysis was of the form:

```
lm(cbind(Cluster1, Cluster2, Cluster3, Cluster4) ~ ., data =
as.data.frame(scale(my_data)))
```

Beyond understanding which roles lead to which outcomes, an additional analysis was completed to further examine and understand the effects of personality and other individual constructs on speech acts directly. In other words, this analysis seeks to determine which individual characteristics lead individuals to use certain speech acts within their given roles. To complete this analysis, we performed an additional regression using the ratio of certain speech acts to total speech acts as dependent variables. Results are listed in Table 5. The 95% confidence intervals of the regression coefficients are shown in Figure 6.

Table 5. Regression with Speech Acts as Dependent Variables

Dep Variable:	%ACCEPT		%CLAIM		%MANAGE		%INFORM	
Ind Variable	β	p	β	p	β	p	β	p
Intercept	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Message Contribution	-0.204**	0.006	-0.069	0.359	0.244**	0.001	-0.128	0.089
Avg. Message Length	-0.141	0.066	0.251**	0.002	0.096	0.217	0.153	0.055
Personality								
Agreeableness	0.246**	0.006	0.003	0.970	0.025	0.783	-0.022	0.810
Conscientious	-0.033	0.706	0.116	0.209	0.072	0.423	-0.183*	0.048
Extraversion	-0.136	0.089	0.056	0.499	-0.089	0.272	0.049	0.556
Neuroticism	-0.018	0.825	-0.059	0.490	0.131	0.116	-0.109	0.202
Openness	0.092	0.242	0.047	0.563	-0.021	0.797	-0.022	0.790
Cognitive Ability	-0.042	0.575	-0.014	0.853	0.007	0.930	-0.033	0.672
Social Sensitivity	-0.064	0.421	-0.049	0.552	0.055	0.491	-0.010	0.906
Age	-0.126	0.107	0.016	0.842	-0.041	0.604	0.107	0.188
Gender (f)	0.066	0.390	0.011	0.885	0.137	0.077	-0.120	0.130
Native English	0.132†	0.078	-0.013	0.870	0.012	0.878	-0.028	0.718
Motivation	-0.120	0.182	0.027	0.770	-0.156	0.087	0.234*	0.012
Mutual Trust	0.121	0.162	-0.099	0.270	0.042	0.628	-0.047	0.600
Adjusted R-squared	0.097		0.031		0.074		0.032	
Multiple R-squared	0.164		0.103		0.142		0.103	

*$p<0.05$; **$p<0.01$; ***$p<0.001$

The R function call used for the regression analysis was of the form:

```
lm(cbind(%ACCEPT, %CLAIM, %MANAGE, %INFORM) ~ ., data=as.data.frame(scale(my_data)))
```

5 DISCUSSION

5.1 Interpretation of Findings for RQ1

Our first major objective was to determine which individual traits are related to the emergence of roles in synchronous CMC decision-making tasks. Our results (Table 4) show that one of the emerging roles (the listener role) was significantly predicted by agreeableness, demonstrating that personality can play an important part in emerging roles in virtual work. Other roles were not predicted by personality traits or cognitive ability.

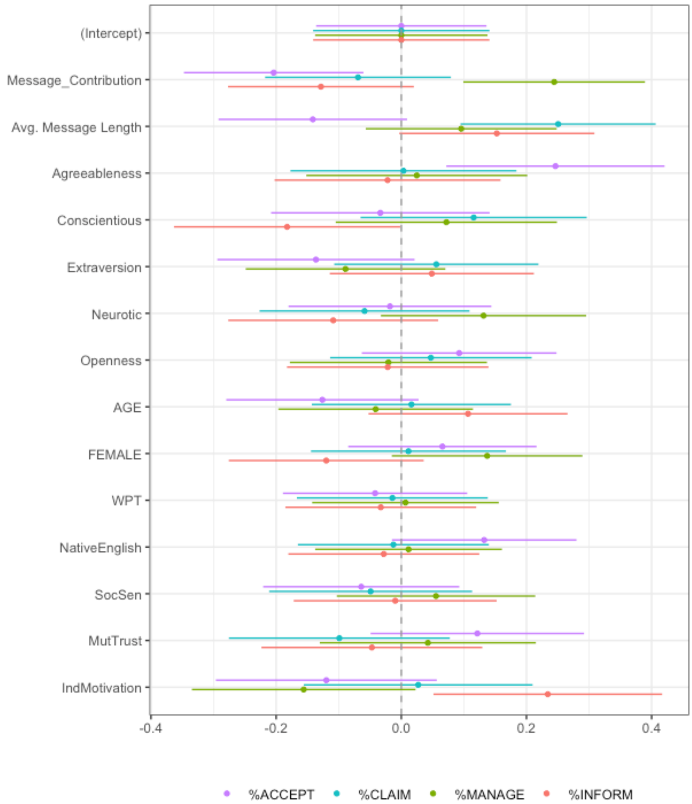

Figure 6. 95% Confidence Intervals for each independent variable with Speech Act as DV

Social sensitivity has a small but significant effect on the extent to which a participant fit the manager role (p=0.014). In other words, those participants who were more socially sensitive were more likely to act as managers. Motivation was also associated with emerging roles, with an increase in motivation on the task leading to higher likelihood of assuming a sharer role. Finally, age was a significant predictor of the sharer role, with older participants more likely to assume this role. We predicted that personality and cognitive ability would affect the emerging roles, and that was true for the listener role, for those higher in agreeableness. However, it also appears that roles emerged independently of these traits. Interestingly, we believe that roles can emerge dynamically, and that the roles presented here may be more likely to emerge based on the interaction of participants and the context of the study than on persistent traits. Finding that persistent traits are not strongly predictive of roles furthers our understanding of the nature of these emergent roles and how they are not always predictable based on the "type of group member" coming into a group interaction.

To further understand the communication, we examined how individual traits directly impact the speech acts used in participation regardless of role. First, we find that personality traits significantly affect participants' usage of accept and inform speech acts. Participants higher in conscientiousness were likely to employ fewer inform speech acts (p=0.048). One unit increase in conscientiousness resulted in using inform speech acts 18 percent less. Agreeableness predicted usage of the accept speech act, with a unit increase of agreeableness leading to a 24% increase in agreeing to others' speech acts.

Second, we find no significant effect of cognitive ability or social sensitivity on the types of speech acts used by participants. We did find that an increase in motivation on the task led to a higher percentage of inform speech acts, meaning that those who were more motivated were more likely to share information. We find it interesting that social sensitivity was predictive of the manager role without being predictive of the MANAGE speech act. This demonstrates that even though the defined roles are each largely based on particular speech acts, the roles do encompass more than the corresponding speech act. For example, although performing a MANAGE speech act (e.g., asking a question or directing the conversation) in isolation does not require particular social sensitivity, those with social sensitivity were more likely to take on this manager role that involved more than simply uttering MANAGE speech acts.

Finally, we found significant relationships between message contributions/length and the types of speech acts used. Higher usage of manage speech acts was positively related with message contribution (p=0.001). The accept speech act was negatively related with message contribution (p=0.006), indicating that those performing accept speech acts tend to be those who share fewer messages. The claim speech act was positively related with message length (p=0.002), meaning that individuals who tend to share longer messages are more likely to make claims in their messages.

Taken altogether, these findings indicate that personality does affect the types of speech acts that participants are likely to use in a group CMC setting. This flows through for agreeableness leading to a listener role; however, personality traits do not predict other emerging roles. The speech acts one uses have some connection to specific persistent traits, whereas the roles assumed, although to a small extent depend on personality, tend to come more dynamically, with the exception of socially sensitive individuals being more likely to take a managing role in the conversation. We conclude that such dynamic role formation is not dependent one's cognitive ability—in other words, the acquisition of roles and the use of speech acts is equal across varying cognitive abilities. One other potential reason that cognitive ability did not predict roles nor speech acts is that cognitive ability typically affects individual performance and is not as strongly predictive of outcomes in collaborative settings, particularly online collaborative settings [56].

5.2 Interpretation of Findings for RQ2

Our second major objective was to replicate the previous Barlow [14, 15] studies. In this study, we validated the results of previous studies and found support for the four roles in synchronous CMC decision-making reported in those studies.

Cluster 1 corresponds to the Sharer role [14], where the main type of communication is informing, and has a higher than average length of messages. This role also tends to share few opinions and the least amount of managing statements. Cluster 2 represents a role where people share mostly opinions (few informing or accepting statements and higher claims). This cluster corresponds with the Opinionater role [14] and has a high message contribution. Cluster 3 corresponds to the Manager role [15], who guides the group conversation by directing actions, requesting information, and asking questions. These Managers share longer messages than any other role (although still not extremely lengthy). They also contribute more messages than others. Cluster 4 corresponds to the Listener role [15], those who contribute the least information in the shortest messages that primarily consist of agreeing with others.

Table 6 provides a summary description, indicating which characteristics are similar to those previously found—these are likely the stable set of characteristics associated with these four roles. An overall synthesis of roles based on all studies is shown in Table 7.

5.3 Limitations

Certain limitations must be considered when interpreting the results. First, this study examined only a decision-making task in a synchronous environment. Consistent with role theory, the same roles will not exist in every context, but some consistencies do emerge. This opens the door for much future research to understand the communication roles that are likely to emerge in various contexts. Further, it would be important in other contexts to understand the individual characteristics that predict roles as this study did.

Table 6. Interpretation of Clusters / Roles

Cluster	Role name	Similar characteristics to roles found by Barlow [14, 15]	Results unique to this study
1	"Sharers"	Large share of messages dedicated to sharing factual information with group members	Even more focused on sharing, with lower amounts of claim and accept speech acts
2	"Opinionaters"	Large share of messages dedicated to making claims, with relatively few messages sharing facts and agreeing	This role is most similar to its respective counterpart in previous studies than any of the other roles
3	"Managers"	Mostly guide the discussion by managing or asking questions; moderate amounts of claims	Fewer instances of agreeing with others
4	"Listeners"	Fewer number of messages, mostly agreeing with others; moderate amount of sharing information	More moderate in guiding discussion

Table 7. Findings about Roles Across Studies

Role name	Feature	Barlow [14]	Barlow [15]	Current Study	Overall
"Sharers"	% Messages	Low	High	Moderate	Varies
	Msg Length	Low	Moderate	Moderate	Moderate leaning low
	% Inform	High	High	High	High
	% Claim	Low	Moderate	Low	Fairly low
	% Accept	Moderate	Low	Low	Fairly low
	% Manage	Low	Moderate	Low	Fairly low
"Opinionaters"	% Messages	High	Moderate	Moderate	Moderately high
	Msg Length	Moderate	High	Moderate	Moderate leaning high
	% Inform	Low	Low	Low	Low
	% Claim	High	High	High	High
	% Accept	Low	Moderate	Low	Fairly low
	% Manage	Moderate	Fairly low	Moderate	Moderate leaning low
"Managers"	% Messages	High	High	Fairly high	Fairly high
	Msg Length	Moderate	Fairly low	Moderate	Varies
	% Inform	Moderate	Low	Moderate	Moderate leaning low
	% Claim	Moderate	Moderate	Fairly low	Moderate
	% Accept	Moderate	Moderate	Low	Moderate leaning low
	% Manage	High	High	High	High
"Listeners"	% Messages	Low	Low	Fairly low	Low
	Msg Length	High	Low	Low	Varies
	% Inform	Moderate	High	Moderate	Moderate leaning high
	% Claim	Moderate	Low	Low	Moderate leaning low
	% Accept	High	High	High	High
	% Manage	Low	Low	Moderate	Moderate leaning low

A second limitation is that the participants in this study were meeting for the first time. Groups who have experience working together, or at least have an understanding of personalities of other group members, may develop different or distinct roles. Thus, this research is most applicable to ad hoc groups or project members working together for the first time. Future research may examine how roles within a group persist or change over time. Further, many organizational groups do have members with pre-defined roles; informal roles, if emerging at all, would likely emerge differently in such groups.

Next, the study was conducted through a lab experiment and has the typical limitations of this method. Participants were students, so care should be taken in generalizing the results to other populations. However, because the phenomenon of interest in this study regards general human behavior, it is anticipated that results should generalize fairly well, as student participants are appropriate for such generalizations [88].

Lastly, the model trained using the LightSide tool achieved a moderate accuracy of 0.72 and a Kappa score of 0.63 for speech act classification. Additionally, the R-squared measures for the roles and speech acts regression models were low, suggesting that there are other explanatory variables that were not accounted for.

5.4 Practical Implications

Understanding the common roles that generally emerge during synchronous CMC interaction can help individuals better anticipate working within groups in this setting by knowing the common types of roles that may emerge in the group. If a person knows the general personality of a teammate, they may be better able to anticipate how that teammate will communicate and react during the course of online decision making. Such knowledge should also help groups and organizations to structure and facilitate group work by better understanding role interaction and how various individuals assume roles.

Understanding the emergent roles of virtual decision-making groups can also inform the design of information technology to support such virtual groups. Systems can be created that facilitate communication and coordination based on communication patterns of individuals within the group, and such interventions may lead to improved group outcomes.

6 CONCLUSION

This study extends research on behavioral roles in computer-mediated communication in several ways. First, this study shows that individual characteristics, both persistent (social sensitivity and agreeableness) and context-specific (motivation), are predictors of assumed communication roles. Yet also interestingly, we find that persistent traits such as personality and cognitive ability do not have as strong effect on the roles that individuals assume as they do on speech acts; rather, these roles likely emerge during interaction in the group in response to other group members and context-specific discussions.

Second, this study was a successful replication of the previous Barlow [14, 15] studies. This validates the types of roles that emerge when groups make task-focused decisions using text-based CMC. Beyond direct replication, future research should continue to examine communication and roles in synchronous CMC decision-making, in other contexts, task types, and media formats.

In addition to looking at the roles themselves, this study also contributes by taking a more nuanced look at the effect of these traits on the speech acts directly. For example, conscientiousness was related with fewer instances of the inform speech act, and agreeableness was associated with higher instances of the accept speech act.

We recognize that individual characteristics cannot alone fully explain the development of roles; role theory posits that roles develop based on a combination of individual characteristics and situational factors. Previous research had examined roles in computer-mediated communication but had not examined what type of individual characteristics would predict these roles. Future research should continue to examine additional individual and group level antecedents of emergent roles and speech patterns in synchronous CMC. Future research should also examine both individual and situational factors together. For example, research could examine role conflict in groups [89] and how individual traits affect the way that such conflict changes the formation of roles.

Third, text mining was used to partially automate speech act analysis as part of the role detection process. The process of automating speech act analysis is fairly recent [90-93]. Few studies examining roles have used automated text analysis to examine message *content* [13, 46, 94]; these studies use text mining to find word patterns largely without focusing on the type of speech act. Future studies can build on this work by utilizing text mining methods to help analyze communication at a large scale to understand roles. For example, future research can apply the framework of role emergence, and the method we employed, to study the transcripts of video meetings, which are much more common now in real-world decision-marking groups. This would allow us to better understand how the factors identified in this study influence role emergence and to develop more effective strategies for promoting collaboration and sharing in these groups. We encourage future research to seek to apply our analysis, coding, and regression models to other settings within the CMC roles.

REFERENCES

[1] Welbourne, T. M., Johnson, D. E. and Erez, A. The role-based performance scale: Validity analysis of a theory-based measure. *Academy of Management Journal*, 41, 5 (1998), 540-555.
[2] Biddle, B. J. *Role Theory: Expectations, Identities, and Behaviors*. Academic Press, 2013.
[3] Kakkar, H., Tangirala, S., Srivastava, N. K. and Kamdar, D. The dispositional antecedents of promotive and prohibitive voice. *Journal of Applied Psychology*, 101, 9 (2016), 1342.
[4] Nordbäck, E. S. and Espinosa, J. A. Effective coordination of shared leadership in global virtual teams. *Journal of Management Information Systems*, 36, 1 (2019), 321-350.
[5] Laitinen, K., Laaksonen, S.-M. and Koivula, M. Slacking with the Bot: Programmable Social Bot in Virtual Team Interaction. *Journal of Computer-Mediated Communication*, 26, 6 (2021), 343-361.
[6] Waizenegger, L., McKenna, B., Cai, W. and Bendz, T. An affordance perspective of team collaboration and enforced working from home during COVID-19. *European Journal of Information Systems*, 29, 4 (2020), 429-442.
[7] Dennis, A. R., Rennecker, J. A. and Hansen, S. Invisible Whispering: Restructuring Collaborative Decision making with Instant Messaging. *Decision Sciences*, 41, 4 (2010), 845-886.
[8] Biddle, B. J. Recent developments in role theory. *Annual Review of Sociology*, 12 (1986), 67-92.
[9] Shaik, F. F. and Makhecha, U. P. Drivers of employee engagement in global virtual teams. *Australasian Journal of Information Systems*, 23 (2019).
[10] Welser, H. T., Gleave, E., Fisher, D. and Smith, M. A. Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, 8, 2 (2007), 1-32.
[11] Arazy, O., Lifshitz-Assaf, H. and Balila, A. Neither a Bazaar nor a cathedral: The interplay between structure and agency in Wikipedia's role system. *Journal of the Association for Information Science and Technology*, 70, 1 (2019), 3-15.
[12] Sun, L., Kraut, R. E. and Yang, D. Multi-level Modeling of Social Roles in Online Micro-lending Platforms. *Proceedings of the ACM on Human-Computer Interaction*, 3, CSCW (2019), 1-25.
[13] Dowell, N. M., Nixon, T. M. and Graesser, A. C. Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior research methods*, 51, 3 (2019), 1007-1041.
[14] Barlow, J. B. Emergent roles in decision-making tasks using group chat. In *Proceedings of the Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (2013). ACM.
[15] Barlow, J. B. Emergent roles in computer-mediated synchronous virtual groups. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (Maui, HI, 2019).
[16] Stewart, G. L., Fulmer, I. S. and Barrick, M. R. An exploration of member roles as a multilevel linking mechanism for individual traits and team outcomes. *Personnel Psychology*, 58, 2 (2005), 343-365.
[17] McEwan, B., Carpenter, C. J. and Westerman, D. On replication in communication science. *Communication Studies*, 69, 3 (2018), 235-241.
[18] Brendel, A. B., Diederich, S. and Niederman, F. An immodest proposal–going “All in” on replication research in information systems. *European Journal of Information Systems* (2021), 1-10.

[19] Akar, E., Mardikyan, S. and Dalgic, T. User roles in online communities and their moderating effect on online community usage intention: An integrated approach. *International Journal of Human–Computer Interaction*, 35, 6 (2019), 495-509.
[20] Gleave, E., Welser, H. T., Lento, T. M. and Smith, M. A. A conceptual and operational definition of 'social role' in online community. In *Proceedings of the Proceedings of the 43rd Hawaii International Conference on System Sciences* (January 5-8, 2009, 2009). Computer Society Press.
[21] Arazy, O., Daxenberger, J., Lifshitz-Assaf, H., Nov, O. and Gurevych, I. Turbulent stability of emergent roles: The dualistic nature of self-organizing knowledge coproduction. *Information Systems Research*, 27, 4 (2016), 792-812.
[22] Kane, G. C., Johnson, J. and Majchrzak, A. Emergent life cycle: The tension between knowledge change and knowledge retention in open online coproduction communities. *Management Science*, 60, 12 (2014), 3026-3048.
[23] Sutanto, J., Phang, C. W., Kuan, H. H., Kankanhalli, A., Cheng, B. and Tan, Y. Vicious and virtuous cycles in global virtual team role coordination. In *Proceedings of the Proceedings of the 38th Hawaii International Conference on System Sciences* (2005). IEEE.
[24] Krishna, S., Billot, R. and Jullien, N. A clustering approach to infer Wikipedia contributors' profile. In *Proceedings of the Proceedings of the 14th International Symposium on Open Collaboration* (2018).
[25] Maki, K., Yoder, M., Jo, Y. and Rosé, C. Roles and success in wikipedia talk pages: Identifying latent patterns of behavior. In *Proceedings of the Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (2017).
[26] Bulgurcu, B., Van Osch, W. and Kane, G. C. The rise of the promoters: user classes and contribution patterns in enterprise social media. *Journal of Management Information Systems*, 35, 2 (2018), 610-646.
[27] Choi, J. O., Herbsleb, J., Hammer, J. and Forlizzi, J. Identity-Based Roles in Rhizomatic Social Justice Movements on Twitter. In *Proceedings of the Proceedings of the International AAAI Conference on Web and Social Media* (2020).
[28] Benamar, L., Balagué, C. and Ghassany, M. The identification and influence of social roles in a social media product community. *Journal of Computer-Mediated Communication*, 22, 6 (2017), 337-362.
[29] Kou, Y., Gray, C., Toombs, A. and Adams, R. Knowledge production and social roles in an online community of emerging occupation: A study of user experience practitioners on Reddit. In *Proceedings of the Proceedings of the 51st Hawaii International Conference on System Sciences* (2018).
[30] Sharma, A., Choudhury, M., Althoff, T. and Sharma, A. Engagement Patterns of Peer-to-Peer Interactions on Mental Health Platforms. In *Proceedings of the Proceedings of the International AAAI Conference on Web and Social Media* (2020).
[31] Turner, T. C., Smith, M. A., Fisher, D. and Welser, H. T. Picturing Usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication*, 10, 4 (2005).
[32] Chang, C.-K., Chen, G.-D. and Wang, C.-Y. Statistical model for predicting roles and effects in learning community. *Behaviour & Information Technology*, 30, 1 (2011), 101-111.
[33] Welser, H. T., Kossinets, G., Smith, M. A. and Cosley, D. Finding social roles in Wikipedia. In *Proceedings of the Annual Meeting of the ASA* (Boston, MA, 2008). AllAcademic.
[34] Liu, J. and Ram, S. Who does what: Collaboration patterns in the Wikipedia and their impact on article quality. *ACM Transactions on Management Information Systems*, 2, 2 (2011), article 11.
[35] Brooks, I. and Swigger, K. Using sentiment analysis to measure the effects of leaders in global software development. In *Proceedings of the 2012 International Conference on Collaboration Technologies and Systems (CTS)* (2012). IEEE.
[36] Chiu, C. H., Yang, H. Y., Liang, T. H. and Chen, H. P. Elementary students' participation style in synchronous online communication and collaboration. *Behaviour & Information Technology*, 29, 6 (2010), 571-586.
[37] Carte, T. A., Chidambaram, L. and Becker, A. Emergent leadership in self-managed virtual teams. *Group Decision and Negotiation*, 15, 4 (2006), 323-343.
[38] Austin, J. L. *How to Do Things with Words*. Harvard University Press, Cambridge, MA, USA, 1962.
[39] Herring, S. C. *Computer-mediated discourse analysis: An approach to researching online behavior*. Cambridge University Press, City, 2004.
[40] Arnold, T., Daxenberger, J., Gurevych, I. and Weihe, K. *Is Interaction More Important Than Individual Performance? A Study of Motifs in Wikia*. City, 2017.
[41] Cheng, J. and Guo, J. L. Activity-based analysis of open source software contributors: Roles and dynamics. In *Proceedings of the 2019 IEEE/ACM 12th International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)* (2019). IEEE.
[42] Frank, L., Gimpel, H., Schmidt, M. and Schoch, M. Emergent User Roles of a Digital Workplace: A Network Analysis Based on Trace Data. In *Proceedings of the International Conference on Information Systems (ICIS)* (2017).
[43] Gimpel, H., Haamann, F., Schoch, M. and Wittich, M. User roles in online political discussions: a typology based on twitter data from the german federal election 2017. In *Proceedings of the European Conference on Information Systems (ECIS)* (2018).
[44] Mukta, S. H., Ali, M. E. and Mahmud, J. *Deriving interpersonal role identities from social network interactions*. City, 2019.
[45] Schwade, F. and Schubert, P. *Developing a User Typology for the Analysis of Participation in Enterprise Collaboration Systems*. City, 2019.
[46] Seering, J., Flores, J. P., Savage, S. and Hammer, J. The social roles of bots: evaluating impact of bots on discussions in online communities. *Proceedings of the ACM on Human-Computer Interaction*, 2, CSCW (2018), 1-29.

[47] Hara, N. and Sanfilippo, M. R. Analysis of roles in engaging contentious online discussions in science. *Journal of the Association for Information Science and Technology*, 68, 8 (2017), 1953-1966.

[48] Yang, D., Kraut, R. E., Smith, T., Mayfield, E. and Jurafsky, D. Seekers, providers, welcomers, and storytellers: Modeling social roles in online health communities. In *Proceedings of the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019).

[49] Morgeson, F. P., Reider, M. H. and Campion, M. A. Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel psychology*, 58, 3 (2005), 583-611.

[50] McCrae, R. R. and John, O. P. An introduction to the five-factor model and its applications. *Journal of Personality*, 60, 2 (1992), 175-215.

[51] Bell, S. T. Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, 92, 3 (2007), 595.

[52] Curşeu, P. L., Ilies, R., Vîrgă, D., Maricuţoiu, L. and Sava, F. A. Personality characteristics that are valued in teams: Not always "more is better"? *International Journal of Psychology*, 2018, 1 (2018).

[53] Peeters, M. A. G., Van Tuijl, H. F., Rutte, C. G. and Reymen, I. M. Personality and team performance: A meta-analysis. *European Journal of Personality*, 20, 5 (2006), 377-396.

[54] Balthazard, P. A., Potter, R. E. and Warren, J. E. The effects of extraversion and expertise on virtual team interaction and performance. In *Proceedings of the Proceedings of the 35th Hawaii International Conference on System Science (HICSS)* (2002). IEEE.

[55] Turel, O. and Zhang, Y. Does virtual team composition matter? Trait and problem-solving configuration effects on team performance. *Behaviour & Information Technology*, 29, 4 (2010), 363-375.

[56] Dennis, A. S., Barlow, J. B. and Dennis, A. R. The power of introverts: Personality and intelligence in virtual teams. *Journal of Management Information Systems*, 39, 1 (2022), 102-129.

[57] Balthazard, P. A., Waldman, D. A. and Warren, J. E. Predictors of the emergence of transformational leadership in virtual decision teams. *The Leadership Quarterly*, 20, 5 (2009), 651-663.

[58] Straus, S. G. Getting a clue: The effects of communication media and information distribution on participation and performance in computer-mediated and face-to-face groups. *Small Group Research*, 27, 1 (1996), 115-142.

[59] Gutiérrez Maldonado, J., Mora, M., García, S. and Edipo, P. Personality, sex and computer-mediated communication through the Internet. *Anuario de PSicologia*, 32, 2 (2001), 51-62.

[60] Deary, I. J. *Looking Down on Human Intelligence: From Psychometrics to the Brain*. Oxford University Press, New York, NY, 2000.

[61] Stewart, G. L. A meta-analytic review of relationships between team design features and team performance. *Journal of Management*, 32 (2006), 29-54.

[62] Schmidt, F. L. and Hunter, J. General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 1 (2004), 162-173.

[63] Jung, J. H., Schneider, C. and Valacich, J. Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science*, 56, 4 (2010), 724-742.

[64] Valacich, J. S., Jung, J. H. and Looney, C. A. The effects of individual cognitive ability and idea stimulation on idea-generation performance. *Group Dynamics: Theory, Research, and Practice*, 10, 1 (2006), 1-15.

[65] Barlow, J. B. and Dennis, A. R. Not as smart as we think: A study of collective intelligence in virtual groups. *Journal of Management Information Systems*, 33, 3 (2016), 684-712.

[66] Paulhus, D. L. and Morgan, K. L. Perceptions of intelligence in leaderless groups: The dynamic effects of shyness and acquaintance. *Journal of Personality and Social Psychology*, 72, 3 (1997), 581-591.

[67] Engel, D., Woolley, A. W., Jing, L. X., Chabris, C. F. and Malone, T. W. Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PloS one*, 9, 12 (2014), e115212.

[68] Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. and Plumb, I. The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger Syndrome or High-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 2 (2001), 241-251.

[69] Bender, L., Walia, G., Kambhampaty, K., Nygard, K. E. and Nygard, T. E. Social sensitivity and classroom team projects: An empirical investigation. In *Proceedings of the SIGCSE 2012 Conference* (Raleigh, NC, 2012).

[70] Hassell, M. D. and Cotton, J. L. Some things are better left unseen: Toward more effective communication and team performance in video-mediated interactions. *Computers in Human Behavior*, 73 (2017), 200-208.

[71] Bartelt, V. L., Dennis, A. R., Yuan, L. I. and Barlow, J. B. Individual priming in virtual team decision-making. *Group Decision and Negotiation* 22, 5 (2013), 873-896.

[72] John, O. P., Naumann, L. P. and Soto, C. J. *Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues*. Guilford Press, City, 2008.

[73] Dodrill, C. B. Long-term reliability of the Wonderlic Personnel Test. *Journal of Consulting and Clinical Psychology*, 51, 2 (1983), 316-317.

[74] Dodrill, C. B. and Warner, M. H. Further studies of the Wonderlic Personnel Test as a brief measure of intelligence. *Journal of Consulting and Clinical Psychology*, 56, 1 (1988), 145-147.

[75] Edmondson, A. Psychological safety and learning behavior in work teams. *Administrative science quarterly*, 44, 2 (1999), 350-383.

[76] Fransen, J., Kirschner, P. A. and Erkens, G. Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness. *Computers in human Behavior*, 27, 3 (2011), 1103-1113.

[77] Azur, M. J., Stuart, E. A., Frangakis, C. and Leaf, P. J. Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20, 1 (2011), 40-49.

[78] Fisher, D., Smith, M. A. and Welser, H. T. You are who you talk to: Detecting roles in Usenet newsgroups. In *Proceedings of the Proceedings of the 39th Hawaii International Conference on System Science (HICSS)* (2006).

[79] Demediuk, S., York, P., Drachen, A., Walker, J. A. and Block, F. Role Identification for Accurate Analysis in Dota 2. In *Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* (2019).

[80] Searle, J. R. *A taxonomy of illocutionary acts.* Cambridge University Press, City, 1979.

[81] Janson, M. A. and Woo, C. C. A speech act lexicon: An alternative use of speech act theory in information systems. *Information Systems Journal*, 6, 4 (1996), 301-329.

[82] Herring, S. C., Das, A. and Penumarthy, S. *CMC act taxonomy*. https://homes.luddy.indiana.edu/herring/cmc.acts.html, 2005.

[83] Kapidzic, S. and Herring, S. C. Gender, communication, and self-presentation in teen chatrooms revisited: Have patterns changed? *Journal of Computer-Mediated Communication*, 17, 1 (2011), 39-59.

[84] Mayfield, E. and Rosé, C. P. *LightSIDE: Open source machine learning for text.* Routledge Academic, City, 2013.

[85] Hartigan, J. A. and Wong, M. A. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28, 1 (1979), 100-108.

[86] Kodinariya, T. M. and Makwana, P. R. Review on determining number of cluster in k-means clustering. *International Journal of Advanced Research in Computer Science and Management Studies*, 1, 6 (2013), 90-95.

[87] Bezdek, J. C., Ehrlich, R. and Full, W. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10, 2/3 (1984), 191-203.

[88] Compeau, D., Marcolin, B., Kelley, H. and Higgins, C. Research commentary—Generalizability of information systems research using student subjects—A reflection on our practices and recommendations for future research. *Information systems research*, 23, 4 (2012), 1093-1109.

[89] Liu, Z., Wang, X., Min, Q. and Li, W. The effect of role conflict on self-disclosure in social network sites: An integrated perspective of boundary regulation and dual process model. *Information Systems Journal*, 29, 2 (2019), 279-316.

[90] Gautam, D., Maharjan, N., Graesser, A. C. and Rus, V. Automated Speech Act Categorization of Chat Utterances in Virtual Internships. In *Proceedings of the International Educational Data Mining Society* (2018).

[91] Wood, A., Rodeghero, P., Armaly, A. and McMillan, C. Detecting speech act types in developer question/answer conversations during bug repair. In *Proceedings of the Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (2018).

[92] Jeong, M., Lin, C.-Y. and Lee, G. G. Semi-supervised speech act recognition in emails and forums. In *Proceedings of the Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (2009).

[93] Morales-Ramirez, I., Kifetew, F. M. and Perini, A. Speech-acts based analysis for requirements discovery from online discussions. *Information Systems*, 86 (2019), 94-112.

[94] Wijenayake, P., Silva, D. d., Alahakoon, D. and Kirigeeganage, S. Automated Detection of Social Roles in Online Communities using Deep Learning. In *Proceedings of the Proceedings of the 3rd International Conference on Software Engineering and Information Management* (2020).

Received July 2023; revised October 2023; accepted November 2023.