Conversational Turn-taking as a Stochastic Process on Networks

Lisa O'Bryan[1], Santiago Segarra[2], Jensine Paoletti[1], Stephanie Zajac[1], Margaret E. Beier[1],
Ashutosh Sabharwal[2], Matthew Wettergreen[3], and Eduardo Salas[1]

Abstract—**Understanding why certain individuals work well (or poorly) together as a team is a key research focus in the psychological and behavioral sciences and a fundamental problem for team-based organizations. Nevertheless, we have a limited ability to predict the social and work-related dynamics that will emerge from a given combination of team members. In this work, we model vocal turn-taking behavior within conversations as a parametric stochastic process on a network composed of the team members. More precisely, we model the dynamic of exchanging the 'speaker token' among team members as a random walk in a graph that is driven by both individual level features and the conversation history. We fit our model to conversational turn-taking data extracted from audio recordings of multinational student teams during undergraduate engineering design internships. Through this real-world data we validate the explanatory power of our model and we unveil statistically significant differences in speaking behaviors between team members of different nationalities.**

I. INTRODUCTION

Data consisting of entities in interconnected systems are ubiquitous in multiple fields. Thus, network structures are commonly used across many disciplines for representation and analysis of complex information [1], from neuroscience [2] to wireless communications [3]. In this work, we represent interactions within teams as small in-person social networks.

Computational models can be an effective way to study the social dynamics that emerge from individuals interacting within groups [4]. In particular, a variety of modeling approaches have been used to try to replicate natural turn-taking behaviors observed in conversation [5]–[8]. Many of these models have been successful in replicating realistic patterns of conversational turn-taking. However, the ability to understand the driving mechanisms underlying these patterns and generalize to novel team compositions is lacking. As a step towards this goal, we develop a stochastic model of conversations that can be used to explore how individual differences impact the emergence of turn-taking patterns within teams. More precisely, we propose a parametric model that captures the individuals' innate tendency to speak as well as the effect that having spoken recently has on speaking again. At every point in time, the next speaker is drawn

from a probability distribution determined by the history of speakers and the aforementioned parameters. The model replicates the majority of conversational turn-taking patterns observed in our real-world data, and our results highlight the important role the memory function plays in replicating these patterns. Furthermore, our results indicate that differences in team member nationality can play a strong role in shaping communication patterns within multinational teams.

Contributions. The contributions of our work are twofold:
i) We propose a simple parametric stochastic process that can capture complex behaviors observed in real data.
ii) We present a novel dataset of conversational turn-taking in undergraduate teams and we apply our model to reveal significant differences in speaking behavior between student nationalities.

II. CONVERSATION MODEL

Inspired by a model by Stasser and Taylor [7], our model incorporates two key notions: i) the relative likelihood π_i that team member i speaks on a given turn independent of their speaking history and ii) the effect m_i that an individual's current speaking turn has on their likelihood of speaking on subsequent turns. We consider the inherent likelihood of speaking π_i of each member i as independent of the history of exchanges and, thus, a constant throughout the conversation. In contrast, we encode dependencies within a conversation through the (turn-dependent) memory function $m_i(t)$. More precisely, for a given member i and a turn t, the memory value is given by

$$m_i(t) = d_i \, e^{-0.5 \, (t - t_i^{\text{last}})}, \tag{1}$$

where d_i is a learnable parameter that controls the scale of the memory effect for each individual and t_i^{last} denotes the last turn on which member i spoke. The negative exponential form in (1) reveals that the memory value asymptotically decreases to 0 as $(t - t_i^{\text{last}})$ increases, i.e., as more turns occur since the last time that member i spoke. This encodes the natural assumption that whether or not an individual spoke many turns ago is inconsequential to their likelihood of speaking next. The memory function in (1) is combined with the innate speaking tendencies π_i to compute the likelihood $\ell_i(t)$ that member i speaks at turn t as follows

$$\ell_i(t) = \begin{cases} 0, & \text{if } t_i^{\text{last}} = t - 1, \\ \pi_i + m_i(t), & \text{otherwise.} \end{cases} \tag{2}$$

Speakers are not allowed to speak on two consecutive turns since these would simply be considered part of the same turn.

This is enforced in (2) by setting the likelihood to zero for the member that has just spoken. Lastly, denoting by N the total number of team members, the likelihoods $\ell_i(t)$ are normalized to sum up to 1 so that they define bona fide probabilities $p_i(t)$ as follows

$$p_i(t) = \frac{\ell_i(t)}{\sum_{j=1}^{N} \ell_j(t)}. \qquad (3)$$

The speaker at turn t is then drawn from this probability distribution across team members.

In summary, the conversational behavior of each individual i within our model is given by two parameters (π_i, d_i). Given these parameters for every team member, the model provides a well-defined stochastic process to generate conversations by the team. More precisely, to determine the speaker at turn t, we first compute the memory values of each member following (1), we then compute likelihoods and transform those into probabilities following (2) and (3), respectively, and we finally draw the next speaker from that probability distribution.

Our main departure from Stasser and Taylor [7] is that our model is based on individual-level parameters whereas theirs is based on team-level parameters. Specifically, Stasser and Taylor's [7] model depends on a single parameter r that determines the inherent speaking probability of every individual (what we denote by π_i) as well as a single parameter d that determines the scale of the memory function for every team member. This fundamental difference is a key enabler for our study of how individual traits relate to each team member's conversational behavior since, given observed conversations of a team, our model enables the estimation of the parameters (π_i, d_i) for every team member.

Given observed turn-taking data, we can fit our model by selecting the parameters (π_i, d_i) for every team member that maximize the probability of generating the observed data. More precisely, if we denote by $\mathcal{H}_{t-1}$ the history of turn-taking up to turn $t - 1$ in the observed data for a given team, and by h_t the speaker at turn t, we can compute the probability that our model selects that true speaker h_t [cf. (3)]. Following the notation in (3), we denote this probability by $p_{h_t}(t \mid \mathcal{H}_{t-1}, \{(\pi_i, d_i)\}_{i=1}^{n})$, i.e., the probability of selecting the true speaker h_t at turn t but where we have now made explicit that this value depends on the past history $\mathcal{H}_{t-1}$ and the parameters (π_i, d_i) for each of the n members in the team. With this notation in place, the log-likelihood of observing the true history of T turns is given by

$$\mathcal{L}(\mathcal{H}_T \mid \{(\pi_i, d_i)\}_{i=1}^{n}) = \sum_{t=1}^{T} \log p_{h_t}(t \mid \mathcal{H}_{t-1}, \{(\pi_i, d_i)\}_{i=1}^{n}). \qquad (4)$$

We fit our model by finding the parameters $\{(\pi_i, d_i)\}_{i=1}^{n}$ that maximize (4). We also fit a *reduced model* that does not contain the memory parameters $\{d_i\}_{i=1}^{n}$ or, equivalently, where $d_i = 0$ for all i. We did this to determine the minimal viable model that can explain our observed data.

III. DATASET

In 2016 and 2017, we collected data on team interactions in student engineering design teams during 7-week internships at a private university in the southern United States. The first week of the internship consisted of a condensed course on the engineering design process, which helped to ensure all participants had a similar baseline level of knowledge. During the remaining six weeks, team members worked together to plan and execute their project which sought to meet a real-world need. We collected data from 7 multi-national teams with team members from the United States (n = 13), Malawi (n = 7), and Brazil (n = 4), with equal numbers of female and male participants. After consenting to the study, participants completed a self-report survey of their personality traits, attitudes, and demographic information. From these data we extracted five features for each individual that we hypothesized could relate to individual differences in speaking patterns, namely, extraversion, agreeableness, conscientiousness, sex (male, female), and nationality (American, Non-American). Our dataset includes multiple meetings from throughout the internships for all teams. Audio streams were processed by annotating the start and end times of speaking turns by each team member during the meetings. Overall, we extracted a mean (SD) of 1941 (1416.5) speaking turns per team.

IV. NUMERICAL EXPERIMENTS

To assess the predictive power of the full and reduced (i.e., without the memory component) models, we perform the following three classes of experiments.

Predicting the next speaker. For each team, we split their turn-taking history into a training and a testing set. The training data contains the first 80 percent of the total turns whereas the testing data contains the remaining turns. As previously explained, we compute the maximum likelihood estimates of the model parameters but this time based only on maximizing the probability of observing the history of the training dataset. We then compute the log-likelihood of observing the history of the testing dataset, as in (4), for both the full and reduced fitted models. The larger (less negative) the attained value, the better predictive power of the corresponding model.

Overall, the full simulation model (i.e., with memory parameter) predicts the observed data better than the reduced simulation model. Table I shows the log-likelihoods attained for the testing dataset (last 20 percent of speaking turns) for each team and simulation model. The full simulation model consistently yields larger (less negative) log-likelihoods, indicating a better predictive performance.

Reproducing relevant conversation patterns.

We test the fit of both the full and reduced simulation models by comparing three measures between our observed and simulated datasets: 1) the proportion of time in which each team member spoke, 2) the proportion of a given speaker's speaking turns following an ABA format in which there was one turn by a different speaker between a given speaker's sequential turns (reflecting the proportion of turns that were part of dyadic exchanges), and 3) the proportion of turns that

TABLE I
LOG-LIKELIHOOD ATTAINED BY BOTH MODELS

Team	No Memory	Memory
Team 1	-159.1586	-154.4533
Team 2	-45.3267	-43.2123
Team 3	-196.5418	-188.6291
Team 4	-278.8524	-255.2545
Team 5	-909.8953	-617.3478
Team 6	-623.9848	-589.0674
Team 7	-105.8199	-104.0864

were part of long dyadic exchanges (4 or more consecutive speaking turns (e.g. ABAB) between two team members). We calculate these measures for each of 10,000 replications of our simulation models and compare them to the values found in our observed data from each team. For each of our three measures, we find the proportion of model replications in which the observed value in the real data fell within the 95 percent confidence interval for values produced by each simulation model. For each of the three speaking patterns of interest, we use chi-squared tests to compare the number of individuals or dyads across teams whose behaviors are not significantly different from those displayed by the full and reduced simulation models.

The full simulation model matches the patterns displayed by significantly more individuals and dyads across teams than the reduced simulation model. The full simulation model correctly estimates the proportion of speaking turns spoken by each team member for 100% (24/24) of team members whereas the reduced simulation model correctly estimates the proportion for 70.8% (17/24) of team members ($\chi^2 = 6.0$, p = 0.014; Figure 1(a)). Moreover, the full simulation model correctly estimates the proportion of each team member's speaking turns with an ABA format for 87.5% (21/24) of team members, but the reduced simulation model correctly estimates the proportion for 29.2% (7/24) of team members, with the tendency to underestimate the proportion of turns ($\chi^2 = 14.5$, p = 0.00014; Figure 1(b)). Finally, the full simulation model correctly estimates the proportion of speaking turns that were part of dyadic exchanges of length 4 turns or greater for 86.7% (26/30) of team member dyads, and the reduced simulation model correctly estimates the proportion for 36.7% (11/30) of team member dyads, with the tendency to underestimate the proportion of turns ($\chi^2 = 13.8$, p = 0.00020; Figure 1(c)).

Relating individual traits and speaking behavior. To gain insight into the relative importance of different individual traits in understanding speaking behaviors, we use an information-theoretic approach [9] to determine which trait(s) best explain between-individual variation in model parameters π_i (baseline likelihood of speaking) and d_i (likelihood of speaking again after speaking recently). Using the MuMIn function [10] in R, we examine which linear model (i.e., a null model and 5 uni-variate models consisting of each of our individual-level predictor variables; see Section III) best explains variation in each parameter value across team members. We group-mean center our three continuous variables (extraversion, agreeableness, conscientiousness) to reflect the relative values of these personality traits among team members. We limit the number

of variables per linear model to one to avoid overfitting. We rank our linear models according to the Akaike information criterion adjusted for small sample sizes (AICc) [9]. We consider top-performing linear models to be the best performing model (i.e. the model with the lowest AICc value) in our model set and any model less than 2 AICc points greater than the best performing linear model [9]. We examine the correlation between individual traits and simulation model parameters for all top-performing linear models to determine how these traits shape speaking behaviors.

Tables II and III display the results of our model selection analysis that compares the relative ability of each of our five univariate models to explain between-individual variation in π_i and d_i. The tables display the AICc values for each model as well as the ΔAICc values (relative to the best model) and corresponding model weights. Model weights reflect the relative support for a given linear model compared to the other candidate models, with 1 indicating full support.

TABLE II
MODEL SELECTION FOR BASELINE LIKELIHOOD OF SPEAKING π_i

model	df	AICc	Δ	wt
Nationality	3	-19.6	0.0	0.94
Null	2	-12.4	7.2	0.025
Agreeableness	3	-11.4	8.2	0.015
Sex	3	-10.3	9.3	0.010
Extraversion	3	-10.0	9.6	0.010
Conscientiousness	3	-9.9	9.7	0.010

TABLE III
MODEL SELECTION FOR SHAPE OF MEMORY FUNCTION d_i

model	df	AICc	Δ	wt
Null	2	105.2	0.0	0.42
Extraversion	3	107.5	2.4	0.13
Nationality	3	107.7	2.5	0.12
Conscientiousness	3	107.8	2.6	0.11
Sex	3	107.8	2.6	0.11
Agreeableness	3	107.8	2.6	0.11

The linear model that best explains between-individual differences in π_i, the parameter reflecting the baseline likelihood of initiating a speaking turn, has nationality as the predictor variable. This linear model has a cumulative model weight of 93.5%. The second best linear model is the null model, which has a ΔAICc value 7.2 higher than the best model (Table II). Since this ΔAICc value is greater than our criterion of ΔAICc = 2 [9], we only consider the linear model with nationality as a predictor variable as a top-performing model within our model set. Overall, this model is supported 37.6 times more strongly (evidence ratio = w_i/w_j = 0.94/0.025 = 37.6) than the null model. When we analyze our top model, we find that Americans have significantly higher likelihoods of initiating speaking turns than non-Americans ($\beta = 0.20$, p < 0.01, Figure 2).

The linear model that best explains between-individual differences in d_i, the change in likelihood of speaking after having just spoken, is the null model. This linear model has a cumulative model weight of 41.6%. The second best linear
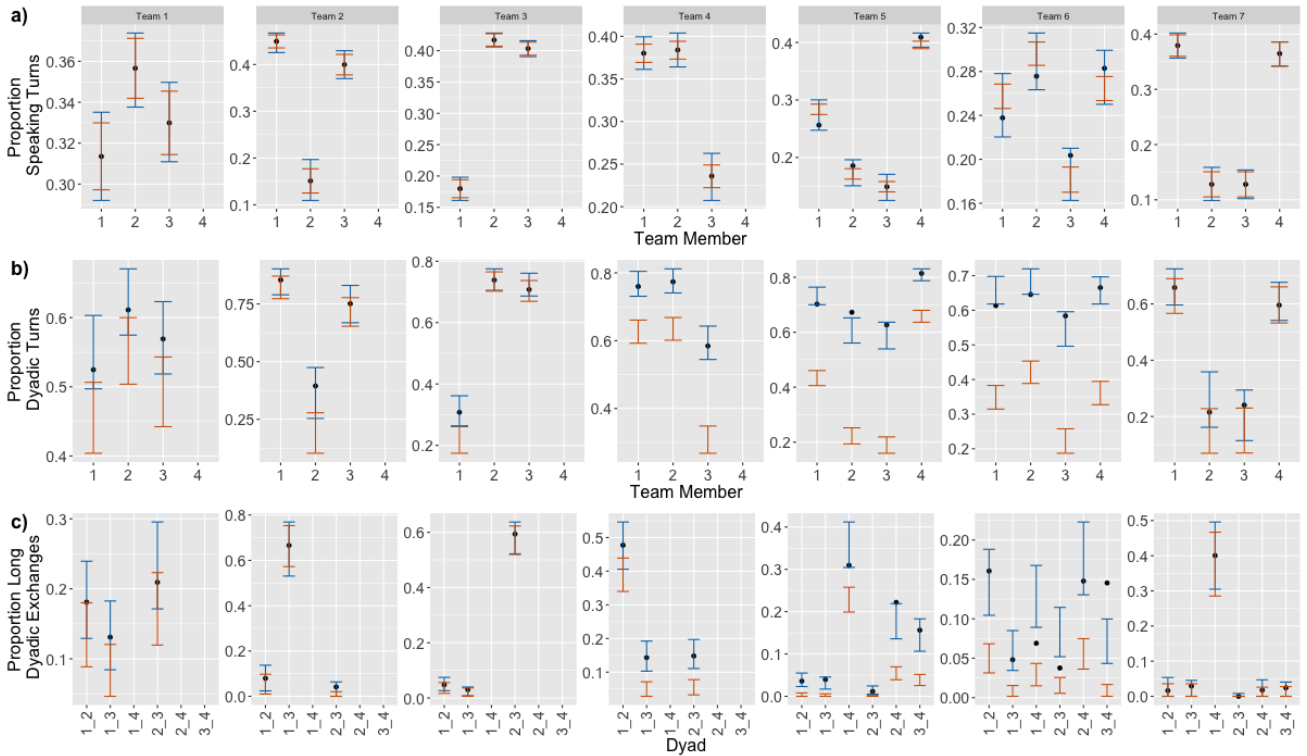
Fig. 1. Black points represent a) observed proportion of speaking turns by each team member, b) observed proportion of speaking turns with one turn in between (e.g. ABA) for each team member, c) observed proportion of speaking turns that were part of consecutive dyadic exchanges of length 4 turns or greater. Error bars represent the 95 percent confidence intervals for the proportions estimated by the reduced simulation model (i.e., without memory parameter) (red) and full simulation model (blue). Y-axis scale varies by team to improve visibility
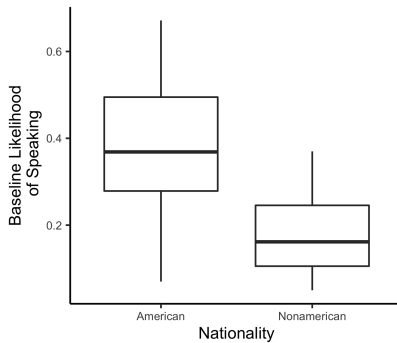

Fig. 2. Boxplot of baseline likelihood of speaking (parameter π_i) by nationality across all teams.

model has extraversion as a predictor variable and a ΔAICc value of 2.4 (Table III). Thus, only the null model is considered a top-performing linear model within our model set, indicating that none of the predictor variables we considered explain between-individual variation in d_i. Overall, the null model is supported approximately 3.2 times more strongly (evidence ratio = w_i/w_j = 0.42/0.13 = 3.2) than the model with extraversion as a predictor.

V. DISCUSSION

The presence of the memory parameter is important in simulating the patterns of vocal turn-taking we observed in our study. Compared to the reduced simulation model with no memory parameter, the full simulation model more accurately predicts future speaking turns and better captures individual and dyadic speaking patterns. This result supports the findings by Stasser and Taylor [7] and Parker [5] that an individual's current likelihood of speaking is impacted by their recent speaking behaviors. Nevertheless, our results differ from those of Stasser and Taylor in that we find different parameter values controlling memory function shape for different individuals.

Our study finds evidence for consistent between-individual differences in speaking behaviors supporting previous findings that individual traits can correlate with communication behaviors [11]–[13]. Our finding that non-Americans initiate speaking turns less frequently than Americans is consistent with a recent study by Li et al. [11] which found that Chinese team members, who tended to be less proficient in English, initiated fewer speaking turns than the American team members. Although we did not measure English language proficiency in our study, our finding could be related to language proficiency since the non-American students in our study were non-native English speakers. Another reason why non-Americans may not have initiated speaking turns as frequently could be that they had a perceived lower status than American team members. Social status may be awarded to the ethnic subgroup with the greatest numerical majority [14]. Since both the Brazilian and Malawian students were completing the internship at

an American university and were outnumbered by American students, they may have demonstrated lower status behaviors like speaking up less frequently [12].

Although nationality best explained differences in baseline frequency of speaking turn initiation, none of our predictor variables explained variation in memory function shape. Future studies are needed to determine whether other traits may explain the observed variation in this speaking tendency. Nevertheless, since Americans were more likely to initiate speaking turns, the broad tendency to speak again after having recently spoken further enhanced individual differences in speaking frequency across team members. Overall, these results help expand knowledge of the impact cultural diversity can have on team processes [15].

A limitation of our study was that we only had data on a relatively small number of teams and team members. This lack of power prevented us from exploring more complex relationships between individual traits and their impacts on speaking behaviors. For example, Neubert and Tagger [16] found that gender moderated the relationship between individual traits and leadership, with certain traits being more important for leadership in males than females and vice versa. Since we tested each of our predictor variables on its own, the strong effect of nationality may have overpowered more subtle or complicated effects of other variables, such as personality and gender. This could be a reason why individual traits like extraversion, which can be strongly correlated with communication tendencies [17]–[19], did not correspond to individual differences in speaking behaviors in our study. Extending our study to more teams would enable a greater understanding of how multiple traits may interact to impact speaking behaviors.

VI. CONCLUSION AND FUTURE WORK

Our study develops a model of conversational turn-taking that can provide a mechanistic understanding of how patterns of communication emerge within teams and can be used to investigate the relationship between team member traits and specific speaking behaviors. Future extensions of our model could integrate more fine-grained speaking behaviors such as the timing between turns and turn overlap, which may enable the study of more complex or subtle turn-taking dynamics. For example, individuals higher in dominance have been found to interrupt more often, which can have a suppressive effect on the speaking behaviors of others [20]. Ultimately, through extensions of our modeling approach, it could be possible to predict the conversational interactions among team members based on their trait composition alone. This ability could enable the anticipation of undesirable team outcomes (e.g., development of subgroups) so that interventions could be applied ahead of time. Similarly, for established teams, it could also be possible to predict the effects team composition changes may have on communication patterns, thus providing guidelines for restaffing or retraining team members. Such predictive models would represent a significant advancement in teams research, enabling a more mechanistic understanding of the connection between team composition and team processes [21], [22].

REFERENCES

[1] I. Brugere, B. Gallagher, and T. Y. Berger-Wolf, "Network structure inference, a survey: Motivations, methods, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 1–39, 2018.

[2] J. D. Medaglia, W. Huang, S. Segarra, C. Olm, J. Gee, M. Grossman, A. Ribeiro, C. T. McMillan, and D. S. Bassett, "Brain network efficiency is influenced by the pathologic source of corticobasal syndrome," *Neurology*, vol. 89, no. 13, pp. 1373–1381, 2017. [Online]. Available: https://n.neurology.org/content/89/13/1373

[3] A. Chowdhury, G. Verma, C. Rao, A. Swami, and S. Segarra, "Unfolding wmmse using graph neural networks for efficient power allocation," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6004–6017, 2021.

[4] J. E. McGrath, H. Arrow, and J. L. Berdahl, "The study of groups: Past, present, and future," *Personality and Social Psychology Review*, vol. 4, no. 1, pp. 95–105, 2000.

[5] K. C. H. Parker, "Speaking Turns in Small Group Interaction: A Context-Free Event Sequence Model," *Journal of Personality and Social Psychology*, vol. 54, no. 6, pp. 965–971, 1988.

[6] S. Basu, T. Choudhury, B. Clarkson, and A. Pentland, "Learning human interactions with the influence model," MIT Media Laboratory, Cambridge, MA, Tech. Rep., 2001.

[7] G. Stasser and L. A. Taylor, "Speaking turns in face-to-face discussions." *Journal of Personality and Social Psychology*, vol. 60, no. 5, pp. 675–684, 1991.

[8] E. Padilha and J. Carletta, "A simulation of small group discussion," in *Workshop on the Semantics and Pragmatics of Dialogue*, 2002, pp. 117–124.

[9] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. Berlin: Springer, 2002.

[10] K. Barton, "MuMIn: Multi-model inference," 2013. [Online]. Available: http://cran.r-project.org/package=MuMIn

[11] H. Li, Y. C. Yuan, N. N. Bazarova, and B. S. Bell, "Talk and let talk: The effects of language proficiency on speaking up and competence perceptions in multinational teams," *Group and Organization Management*, vol. 44, no. 5, pp. 953–989, 2019.

[12] M. S. Mast, "Dominance as expressed and inferred through speaking time," *Human Communication Research*, vol. 28, no. 3, pp. 420–450, 2002.

[13] C. E. Kimble and J. I. Musgrove, "Dominance in arguing mixed-sex dyads: Visual dominance patterns, talking time, and speech loudness," *Journal of Research in Personality*, vol. 22, no. 1, pp. 1–16, 1988.

[14] A. D. Bellmore, A. Nishina, M. R. Witkow, S. Graham, and J. Juvonen, "The influence of classroom ethnic composition on same- and other-ethnicity peer nominations in middle school," *Social Development*, vol. 16, no. 4, pp. 720–740, 2007.

[15] G. K. Stahl, M. L. Maznevski, A. Voigt, and K. Jonsen, "Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups," *Journal of International Business Studies*, vol. 41, no. 4, pp. 690–709, 2010.

[16] M. J. Neubert and S. Taggar, "Pathways to informal leadership: The moderating role of gender on the relationship of individual differences and team member network centrality to informal leadership emergence," *Leadership Quarterly*, vol. 15, no. 2, pp. 175–194, 2004.

[17] K. C. McLean and M. Pasupathi, "Collaborative narration of the past and extraversion," *Journal of Research in Personality*, vol. 40, no. 6, pp. 1219–1231, 2006.

[18] S. K. Leung and M. H. Bond, "Interpersonal communication and personality: Self and other perspectives," *Asian Journal of Social Psychology*, vol. 4, no. 1, pp. 69–86, 2001.

[19] G. A. Macht, D. A. Nembhard, J. H. Kim, and L. Rothrock, "Structural models of extraversion, communication, and team performance," *International Journal of Industrial Ergonomics*, vol. 44, no. 1, pp. 82–91, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.ergon.2013.10.007

[20] E. M. Rogers and D. K. Bhowmik, "Homophily-heterophily: Relational concepts for communication research," *Public Opinion Quarterly*, vol. 34, no. 4, pp. 523–538, 1970.

[21] S. T. Bell, S. G. Brown, A. Colaneri, and N. Outland, "Team composition and the ABCs of teamwork," *American Psychologist*, vol. 73, no. 4, pp. 349–362, 2018.

[22] L. O'Bryan, M. Beier, and E. Salas, "How approaches to animal swarm intelligence can improve the study of collective intelligence in human teams," *Journal of Intelligence*, vol. 8, no. 1, 2020.