

Predicting Group Performances using a Personality Composite-Network Architecture during Collaborative Task

Shun-Chang Zhong^{1,3}, Yun-Shao Lin^{1,3}, Chun-Min Chang^{1,3}, Yi-Ching Liu², Chi-Chun Lee^{1,3}

¹Department of Electrical Engineering, National Tsing Hua University, Taiwan

²College of Management, National Taiwan University, Taiwan

³MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan

flank03200@gmail.com, astanley18074@gmail.com, cmchang@gapp.nthu.edu.tw,
yichingliu@ntu.edu.tw, clee@ee.nthu.edu.tw

Abstract

Personality has not only been studied at an individual level, its composite effect between team members has also been indicated to be related to the overall group performance. In this work, we propose a Personality Composite-Network (P-CompN) architecture that models the group-level personality composition with its intertwining effect being integrated into the network modeling of team members vocal behaviors in order to predict the group performances during collaborative problem solving tasks. In specific, we evaluate our proposed P-CompN in a large-scale dataset consist of three-person small group interactions. Our framework achieves a promising group performance classification accuracy of 70.0%, which outperforms baseline model of using only vocal behaviors without personality attributes by 14.4% absolutely. Our analysis further indicates that our proposed personality composite network impacts the vocal behavior models more significantly on the high performing groups versus the low performing groups.

Index Terms: group interaction, personality traits, attention mechanism, social signal processing

1. Introduction

Small group, which includes three to six people, is the most common composite unit in forming a decision in workplaces. Group scholars have been trying to understand what are the *right* ingredients of the group members that would lead to a better (effective) decision-making process collectively. Group composition is the configuration of member attributes in a team, and the composition of personality has been indicated to have a direct effect on the group performance. In 2007, Bell's meta study shows that each of the Big-5 personality traits can directly impact team's performance [1]. Furthermore, studies have shown that it is more than the individual team member's personality that has an effect, the various configuration of personality attributes between members would bring a different impact to the group [2, 3]. For example, team-level average of both 'openness to experience' and 'emotional stability' moderate the relationship between team conflict and team performance [4]; the variability on 'agreeableness' and 'neuroticism' are negatively related to the team's oral presentation performance [5].

The *right* composition of team members is not only evident in their group-level personality composition but also manifested in 'how each interacts with one another behaviorally' during a small group interaction. The behavioral patterns observed at the group-level during each interaction session are uniquely formed through time as each individual member expresses oneself, exchanges ideas, gears toward consensus or conflictual sit-

uations [6, 7, 8, 9]. In fact, engineering researchers have already made extensive effort into computationally understanding these different interaction processes through automated analyses of verbal and non-verbal behaviors. For example, Okada et al. computed co-occurrences of non-verbal behaviors of each participant with one another to model the group impressions [10]; Fang et al. designed intra- and inter-personal audio-video behavior features to perform personality classification during small group interactions [11]; Batrinca et al. recognized personality attributes using acoustic and visual nonverbal features [12]. Most recently, Lin et al. proposed an interlocutor modulated attention network to reach the state-of-the-art personality recognition accuracy in small group interaction[13].

In this work, we focus on automatically predict group-level performance during a three-person interaction on a collaborative school policy task. There is very limited research on predicting group performance in the past. For example, Murray et al. developed hand-crafted multimodal behavior features of small group conversation to predict team performance on a collaborative task [14]; Avci et al. computed a large set of features including nonverbal multimodal cues, personality traits, diverse interpersonal perception to predict group performances [15]. While Avci et al. integrated personality attributes to demonstrate an improved accuracy, their framework modeled personality attribute simply as auxiliary independent inputs without considering the intertwining effect of group composite personality with individual members behaviors. In fact, group personality is known to influence team performances in two ways, i.e., as an input factor that can increase or decrease the group's overall resources and a modulating factor that shapes teamwork processes [16]. Developing sophisticated frameworks that model the intertwining effect is crucial in advancing the automated behavior modeling in small group interactions.

In this work, we propose a Personality Composite-Network (P-CompN) architecture to predict group performance on a large-scale dataset including 97 sessions of three-person interactions. The P-CompN includes a fusion of two major networks, i.e., Interlocutor Acoustic Network (IAN) and Personality Network (PN). The PN network predicts team performance based on the group-composite Big-5 personality attributes. The IAN network is trained using bi-directional long short-term memory network (BLSTM) with attention being modulated by the group-composite personality. Our P-CompN considers both group-level personality configuration and team member's acoustic behaviors with their intertwining effect being jointly modeled. P-CompN achieves a promising unweighted average recall (UAR) of 70.0% in classifying group performances. Our analysis further reveals that group-composite personality

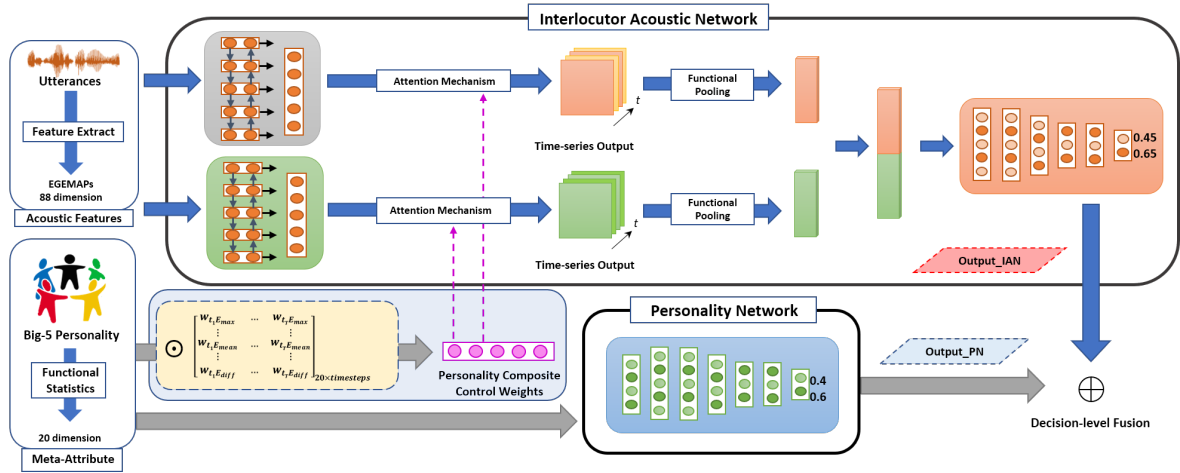


Figure 1: A complete schematic of our Personality Composite-Network (P-CompN). It includes an Interlocutor Acoustic Network (IAN) and a Personality Network (PN) with a decision-level fusion for the classification task. Specifically, we propose to learn a personality composite control weight to modify the original BLSTMs attention mechanism that models the effect of group-level personality attributes on participants acoustic behaviors jointly.

attributes alters significantly the IAN’s attention weights between the high versus the low performing groups.

2. Research Methodology

2.1. The NTHULP Audio-Video Database

Our NTHULP audio-video database is collected at the College of Management of the National Taiwan University (NTU). Each recording includes a session of three participants engaged in a collaborative school policy task [17]. The three participants play different roles chosen at random: vice president of university, vice president of business school, and a member of the business school teachers committee. They are asked to carry out a task to solve school problems by discussing potential alteration on the school policy. Each of the participants would be given a piece of relevant information that is different from others in the team, and they are asked to work together by sharing ideas and communicating collaboratively. However, one of the three participants is a sleeper cell assigned by the experimental personnel. While the sleeper cell knows about all the detailed information to complete the task, he/she would only take part in the task passively. The goal of this task is to study the interaction of the other two participants to understand how they may be influenced by the sleeper cell’s unresponsive behaviors and its effect on the outcome of this collaborative task.

The NTHULP contains 97 recorded sessions with 194 subjects total (age ranges from 19 to 51 years old, 95 males and 99 females). It includes audio and video recordings collected using two cameras and three separate wireless lapel microphones. Additionally, the database contains the following metadata: individual personality trait and group performance outcome score.

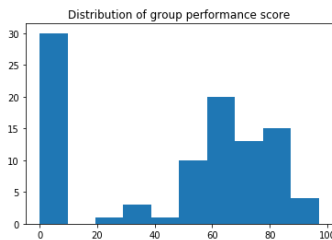


Figure 2: A histogram of the group performance score.

Personality. Each participant’s Big-5 personality attribute, i.e., Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness, is measured using the Goldbergs (1992) 10-item scale [18]. Participants are asked to evaluate how accurately each statement described to them on a 5-point scale, with anchors of 1 = “very inaccurate” and 5 = “very accurate”.

Group Performance. The performance of each team is evaluated by two trained research assistants using the scoring manual for the school policy task developed by Wheeler and Menneck [17]. The scoring manual includes over 300 possible solution scores to this task scenario. The scoring includes two distinct dimensions: a problem-solving score for how well the solution solves the case problem, and a feasibility score for how feasible the solution is to the case problem. The two research assistants independently code all of the 97 groups by identifying the best match between the participants final decision and the solution listed in the manual. Any disagreement between the two coders is reconciled by the third coder.

In this work we use the binarized feasibility score as the class label indicating group performance. We define class 1 as high performing groups with a score greater or equal to 50, and class 0 as low performing groups with a score less than 50. Figure 2 depicts the database distribution of the feasibility score.

2.2. Personality Composite-Network (P-CompN)

Figure 1 shows our Personality Composite-Network (P-CompN) architecture. We model only the two actual participants within the session ignoring the sleeper cell’s behaviors due to his/her consistent non-engaging behaviors in this group performance prediction task. Specifically, our proposed P-CompN architecture is composed of a fusion between two sub-networks, Interlocutor Attention Network (IAN) and Personality Network (PN). We will first describe the two different features inputs to P-CompN and then the details of our framework.

2.2.1. Feature Inputs: Acoustics and Personality Attributes

The audio signals are first segmented into speaker utterances automatically. We extract the extended Geneva minimalistic acoustic parameter set (eGeMAPS) for each utterance [19] as acoustic inputs. eGeMAPS computes 88 dimensional features

Table 1: Model performances using the metric of unweighted average recall (UAR) for high and low performing groups. The overall result shows that the P-CompN outperforms all other methods in group performance classification task achieving 70.0% UAR.

Model (talkative/less)	Individual Models			Group Models			
	Overall	Low	High	Models	Overall	Low	High
Model 0	55.6/54.4	47.2/41.7	64.0/67.2	Model 3	58.1	55.6	60.7
Model 1	55.6/53.2	47.2/50.0	63.9/56.5	Model 4	60.1	61.1	59.0
Model 2	58.5/59.1	69.4/44.4	47.5/73.8	IAN	63.1	63.9	62.3
				PN	63.6	58.3	68.5
				P-CompN	70.0	77.8	62.3

including statistical properties of mel-frequency cepstral coefficients (MFCCs), associated delta, and prosodic information.

In terms of personality attributes, since each of the interlocutors has different traits, in order to measure personality composition characteristics within the group, an intuitive manner is to compute statistics. Specifically, each member has 5 personality scores, and we compute the maximum, minimum, mean and difference value of the group members to derive a 20-dimensional features as inputs of personality attributes.

2.2.2. Interlocutor Acoustic Network (IAN)

The core of IAN uses BLSTM with an attention mechanism trained on the acoustic inputs. Each utterance is a time step t . For each session, we first assign the interlocutors as either a *talkative* or a *talk-less* subject; the talkative subject is the person that speaks the most and often takes the leading role in the interaction, and talk-less subject tends to look quieter and much more tolerant to the existence of an assertive person. We train a typical BLSTM for each subject with attention weight, α_t , defined as:

$$\alpha_t = \frac{\exp(u^T y_t)}{\sum_t \exp(u^T y_t)} \quad (1)$$

where y_t is the hidden layer of time step t .

In this work, we design a novel *personality control mechanism* that integrates the effect of group personality composition into the attention weight. Specifically, we take the 20 dimensional personality composite features multiplies by a learnable weight matrix $W_{20 \times \tau}$ to derive the *personality control weight* for the i -th sample as below:

$$ctrl_{i \times t} = P_{i \times 20} \times W_{20 \times t} \quad (2)$$

where $P_{i \times 20}$ indicates the group composite personality inputs mentioned in section 2.2.1, W is normalized for summing to 1 by softmax. We can then reweight the original attention weight:

$$\alpha'_t = \alpha_t + ctrl_t \quad (3)$$

With this personality reweighted attention mechanism, we further derive the representation of the IAN, z_{IAN} , by concatenating the BLSTM hidden layer output from both the *talkative* and the *talk-less* subject :

$$z_{IAN} = [z''_{talkative} \ z''_{talk-less}] \quad (4)$$

$$z''_{\{talkative, talk-less\}} = G(\alpha'_t \times y_{\{talkative, talk-less\}, t}) \quad (5)$$

where G indicates a *functional pooling layer* over time, i.e., computing the maximum, minimum, mean, median, standard deviation of the hidden layer output for the BLSTM. After obtaining z_{IAN} , we feed it into the prediction layer consists of five fully-connected (DNN) layers to perform binary classification.

2.2.3. Personality Network (PN)

Another sub-network is the personality network (PN). PN is based on a 8-layer DNN that takes the input of 20 dimensional group composite personality features to predict the group performance directly.

The final prediction using our P-CompN in the binary group performances is based on the average softmax output probability of IAN and PN.

3. Experiment Setup and Results

3.1. Experiment Setup

In this section we briefly describe different comparison methods, model parameters, and our evaluation scheme.

3.1.1. Model Comparison

- **Model 0-Baseline**
Using a standard talkative-only or talk-less-only subject's BLSTM with attention (without DNN layers) to perform recognition directly.
- **Model 1-Individual Personality Network**
Using a 8-layer DNN to model talkative or talk-less subject's five personality attributes only (not the composite statistic measures) to perform recognition directly.
- **Model 2-BLSTM + Individual Personality Network**
Combining Model 0 and Model 1 using decision-level fusion by averaging the output probability to perform recognition.
- **Model 3-Dual-BLSTM**
Concatenating output of each interlocutor's BLSTM using a summation pooling layer (not the functional pooling layer) and feeding it to a five-layer DNN to perform recognition.
- **Model 4-Dual-BLSTM + Personality Control**
Integrating Model 3 with the personality control mechanism to the BLSTM attention weight to perform recognition.
- **Interlocutor Acoustic Network (IAN)**
Using the method detailed in section 2.2.2, which modifies Model 4 by replacing summation layer with the functional pooling layer to perform recognition.
- **Personality Network (PN)**
Using the method in section 2.2.3, which uses 5-layer DNN on personality composite features to perform recognition.
- **Personality Composite-Network (P-CompN)**
Using our proposed architecture to perform recognition.

3.1.2. Other Experimental Parameters

We pad sentences to equal length before training (224/147 time-steps for talkative/talk-less respectively), then each BLSTM is trained with a fixed length step. The number of hidden nodes in the BLSTM is 64. IAN has 5 fully-connected layers with node

size of: 1280, 640, 256, 256, 128, 2. PN has 8 fully-connected layers with node size of: 20, 64, 64, 64, 32, 32, 32, 16, 2. We use ReLU as activation function, drop out layer for first and last layers, and batch normalization is also applied. Batch size is set at 16, learning rate is set at 0.0005 using ADAM optimizer. Cross-entropy is our optimized loss function, and we train our network using 40 epochs. The experiment is carried out using 5-folds cross validation using the metric of unweighted average recall (UAR). We adjusted to make the distribution of 5 folds data consistent and reduce the bias.

3.2. Results and Analyses

3.2.1. Analysis on Model Performance

Table 1 summarized our complete prediction results. Our proposed P-CompN obtains the best overall UAR (70.0%), which is 15% higher than baseline Model 0. Model 0 and Model 1 models acoustic behavior and personality attribute using individual participant only (talkative-only or talk-less-only). The accuracy obtained with these two models are only around 55%, and by using complementary information from individual model of acoustic behaviors and personality attribute, i.e., Model 3, it increases slightly to around 59%. We observe that by simply modeling a single participant within a small group collaborative task is not sufficient to obtain a sufficient predictive power of the group performance. Generally, by comparing model accuracy obtained in the ‘Group Models’, i.e., modeling both participants, is better than the ‘Individual Models’.

Furthermore, Model 3 and Model 4 differs by whether the participants acoustic BLSTMs have attention mechanism being modulated by a personality composite control weight. Model 4 improves about 2% over Model 3 in predicting group performance, which indicates that indeed the group personality information affects jointly the behavior manifestation when completing this collaborative task. IAN replaces the conventional summation part of BLSTM attention mechanism with a functional pooling layer, this method computes statistical properties on the time-series output of the BLSTM weighted by personality-controlled attention mechanism. The functional pooling provides another 3% improvement indicating the need of a more complex temporal dynamics characterization of the participants acoustic behaviors, which are shown to be beneficial in this group performance recognition task.

Finally, we also note an interesting observation that PN by itself achieves 63.6% UAR in group performance prediction task. Our experiments demonstrate that the group member’s personality configuration carry significant information on the team performance, which corroborates with past literature in group studies [2, 3]. In summary, our P-CompN architecture that fuses the prediction output of IAN (63.1%) and PN (63.6%) to obtain the best performing model of 70% UAR.

3.2.2. Analysis of Personality and Group Performance

Our experiments demonstrate that personality composite features computed within a group can be used to predict team performance. Table 2: *The bolded number indicates a statistically significant correlation between group performance and each of the Big-5 personality composite attribute.*

Big-5	Max	Min	Mean	Difference
Extraversion	-0.06	-0.04	-0.06	0.07
Agreeableness	0.17	0.06	0.16	0.09
Conscientiousness	-0.01	0.01	-0.01	0.01
Neuroticism	-0.01	0.20	0.08	-0.14
Openness	0.10	-0.05	0.01	0.05

formance classification in this school policy collaborative task. To understand the influence of group personality on team performance, we compute spearman correlation between each of the 20 dimensions of composite group personality measures and our target group performance label.

Table 2 includes the correlation results. The number in bold indicates significant correlation at $\alpha = 0.05$ level. We observe that the maximum, the average of Agreeableness and the minimum of Neuroticism are positively correlated with group performance. Previous study has also shown that Agreeableness is one of the most important personality traits for team performance due to its emphasis on cooperation and facilitation, if the group members could treat others more friendly (max_{Agree}), show patience (min_{Neur}), and keep a collaborative atmosphere ($mean_{Agree}$), it could facilitate a more engaging and comfortable interaction and help finish the task collaboratively with quality [20, 21, 22, 23].

3.2.3. Analysis of Attention Weights

In section 3.2.1, we demonstrate that personality controlled reweighting of the BLSTM attention network help improve the overall prediction accuracy. We would like to further analyze this modified attention weights, which help re-emphasize the important interlocutors behavior regions in the session, as a function of the group performances. Specifically, we compute the ratio within each session that these modified attention weights have positive values, and compare the ratios between the high performing groups versus the low performing groups using t -test ($\alpha = 0.05$). We find that the high performing group sessions have a larger percentage of weights being positive than the low performing group ($p = 0.015$). Our personality controlled weights operates by shifting up and down the original attention weights. Personality mechanism tends to add more weights to the high performing group’s behavior segments. This results seems to be intuitive that for those groups that have the *right* composition of personality configuration would behave more collaboratively, e.g., willing to communicate more, share more ideas, and be more engaging. This is evident in a having larger attention weights placed on their behaviors.

4. Conclusion and Future Work

Personality attribute is not only related to individual behavior pattern during interaction, the personality composition within the group also affects the overall team performance, especially in small group collaborative task solving interactions. In this work, we propose a novel Personality Composite-Network (P-CompN), which includes a personality network (PN) and an interlocutor acoustic network (IAN) that jointly integrate the effect of group members personality attributes into the attention mechanism. We evaluate our P-CompN on a large three-person interaction of school policy task and achieve a promising 70% accuracy in predicting the group performance. Our analyses reveal several important personality attribute configurations to the group performance and demonstrate the effect of higher emphasis on behaviors for groups with higher collaborative effort. We will continue to advance our technical framework by including other non-verbal modalities (e.g., facial expressions and gestures), linguistic contents, and conversation flow (e.g., question answering patterns). Furthermore, by continuously collaborating with group scholars, we would like to investigate the complex interaction effect between the behaviors expressed and the personality traits at the *group*-level and bring insights about the specific interaction strategy that can help better achieve effective communication within a group discussion.

5. References

- [1] S. T. Bell, "Deep-level composition variables as predictors of team performance: a meta-analysis." *Journal of applied psychology*, vol. 92, no. 3, p. 595, 2007.
- [2] A. Kramer, D. P. Bhave, and T. D. Johnson, "Personality and group performance: The importance of personality composition and work tasks," *Personality and Individual Differences*, vol. 58, pp. 132–137, 2014.
- [3] R. L. Moreland, J. Levine, and M. Wingert, "Creating the ideal group: Composition effects at work," *Understanding group behavior*, vol. 2, pp. 11–35, 2013.
- [4] B. H. Bradley, A. C. Klotz, B. E. Postlethwaite, and K. G. Brown, "Ready to rumble: How team personality composition and task conflict interact to improve performance." *Journal of Applied Psychology*, vol. 98, no. 2, p. 385, 2013.
- [5] S. Mohammed and L. C. Angell, "Personality heterogeneity in teams: Which differences make a difference for team performance?" *Small group research*, vol. 34, no. 6, pp. 651–677, 2003.
- [6] K. A. Jehn and E. A. Mannix, "The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance," *Academy of management journal*, vol. 44, no. 2, pp. 238–251, 2001.
- [7] C. Beyan, V.-M. Katsageorgiou, and V. Murino, "A sequential data analysis approach to detect emergent leaders in small groups," *IEEE Transactions on Multimedia*, 2019.
- [8] P. Dhani and T. Sharma, "Emotional intelligence and personality traits as predictors of job performance of it employees," *International Journal of Human Capital and Information Technology Professionals (IJHCITP)*, vol. 9, no. 3, pp. 70–83, 2018.
- [9] N. Attia, *Big Five personality factors and individual performance*. Université du Québec à Chicoutimi, 2013.
- [10] S. Okada, L. S. Nguyen, O. Aran, and D. Gatica-Perez, "Modeling dyadic and group impressions with intermodal and interperson features," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 15, no. 1s, p. 13, 2019.
- [11] S. Fang, C. Achard, and S. Dubuisson, "Personality classification and behaviour interpretation: An approach based on feature categories," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 2016, pp. 225–232.
- [12] L. Batrinca, N. Mana, B. Lepri, N. Sebe, and F. Pianesi, "Multimodal personality recognition in collaborative goal-oriented tasks," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 659–673, 2016.
- [13] Y.-S. Lin and C.-C. Lee, "Using interlocutor-modulated attention blstm to predict personality traits in small group interaction," in *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 2018, pp. 163–169.
- [14] G. Murray and C. Oertel, "Predicting group performance in task-based interaction," in *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ACM, 2018, pp. 14–20.
- [15] U. Avci and O. Aran, "Predicting the performance in decision-making tasks: From individual cues to group interaction," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 643–658, 2016.
- [16] J. E. Driskell, R. Hogan, and E. Salas, *Personality and group performance*. Sage Publications, Inc, 1987.
- [17] B. Wheeler and B. Mennecke, "The school of business policy task manual," 1992.
- [18] L. R. Goldberg, "The development of markers for the big-five factor structure." *Psychological assessment*, vol. 4, no. 1, p. 26, 1992.
- [19] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.
- [20] W. G. Graziano, E. C. Hair, and J. F. Finch, "Competitiveness mediates the link between personality and group performance." *Journal of Personality and Social Psychology*, vol. 73, no. 6, p. 1394, 1997.
- [21] G. A. Van Kleef, A. C. Homan, B. Beersma, and D. van Knippenberg, "On angry leaders and agreeable followers: How leaders emotions and followers personalities shape motivation and team performance," *Psychological Science*, vol. 21, no. 12, pp. 1827–1834, 2010.
- [22] T. Sy, S. Côté, and R. Saavedra, "The contagious leader: impact of the leader's mood on the mood of group members, group affective tone, and group processes." *Journal of applied psychology*, vol. 90, no. 2, p. 295, 2005.
- [23] R. E. De Vries, B. Van den Hooff, and J. A. De Ridder, "Explaining knowledge sharing: The role of team communication styles, job satisfaction, and performance beliefs," *Communication research*, vol. 33, no. 2, pp. 115–135, 2006.