An Exploratory Computational Study on the Effect of Emergent Leadership on Social and Task Cohesion

Soumaya Sabry
soumaya.sabry@telecom-paris.fr
LTCI, Télécom Paris, Institut polytechnique de Paris
91120 Palaiseau, France

Lucien Maman
lucien.maman@telecom-paris.fr
LTCI, Télécom Paris, Institut polytechnique de Paris
91120 Palaiseau, France

Giovanna Varni
giovanna.varni@telecom-paris.fr
LTCI, Télécom Paris, Institut polytechnique de Paris
91120 Palaiseau, France

ABSTRACT

Leadership is a complex and dynamic phenomenon that has received a lot of attention from psychologists over the last 50 years, primarily due to its relationships with team effectiveness and performances. Depending on the group (e.g., size, relationships among members) and the context (e.g., solving a task under pressure), various styles of leadership could emerge. These styles can either be formally decided or manifest informally. Among the informal types of leadership, emergent leadership is one of the most studied. It is an emergent state that develops over time in a group and that interplays with other emergent states such as cohesion. Only a few computational studies focusing on predicting emergent leadership take advantage of the relationships with other phenomena to improve their models' performances. These approaches, however, only apply to their models aimed at predicting emergent leadership. There is, to the best of our knowledge, no approach that integrates emergent leadership into computational models of cohesion.

In this study, we take a first step towards bridging this gap by introducing 2 families of approaches inspired by Social Sciences' insights to integrate emergent leadership into computational models of cohesion. The first family consists of amplifying the differences between leaders' and followers' features while the second one focuses on adding leadership representation directly into the computational model's architecture. In particular, for each family, we describe 2 approaches that are applied to a Deep Neural Network model aimed at predicting the dynamics of cohesion across various tasks over time. This study explores whether and how applying our approaches improves the prediction of the dynamics of the Social and Task dimensions of cohesion. Therefore, the performance of a computational model of cohesion that does not integrate the interplay between cohesion and emergent leadership is compared with the same computational models that apply our approaches. Results show that approaches from both families significantly improved the prediction of the Task cohesion dynamics, confirming the benefits of integrating emergent leadership following Social Psychology's insights to enforce computational models of cohesion at both feature and architecture levels.

ICMI '21 Companion, October 18–22, 2021, Montréal, QC, Canada

CCS CONCEPTS

• **Human-centered computing** → *Collaborative and social computing*; • **Computing methodologies** → *Artificial intelligence*.

KEYWORDS

Cohesion, Emergent Leadership, Multimodal Interaction, Social Signal Processing

ACM Reference Format:
Soumaya Sabry, Lucien Maman, and Giovanna Varni. 2021. An Exploratory Computational Study on the Effect of Emergent Leadership on Social and Task Cohesion. In *Companion Publication of the 2021 International Conference on Multimodal Interaction (ICMI '21 Companion), October 18–22, 2021, Montréal, QC, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3461615.3485415

1 INTRODUCTION

"*The greatest leader is not necessarily the one who does the greatest things. He is the one that gets the people to do the greatest things*". In his famous interview for CBS, Ronald Reagan points out an important feature of a leader. Leadership is not only reflecting someone's ability to perform a task or to give instructions but also its capacity to inspire and lead others to achieve their goal. Given the complex nature of leadership, scholars in Social Psychology suggest various definitions (e.g., [14]) and describe leadership as an emergent state [2]. Emergent states are usually group phenomena that result from the micro-level affective, behavioral, and cognitive interactions among group members during an interaction (e.g., [28]). Emergent leadership, however, is an emergent state that manifests at an individual level. It characterizes a person who appears as the leader during social interaction, without any formal authority [48]. Existing studies on automated emergent leadership detection in small groups showed reasonable performance using nonverbal features (e.g., [6, 8, 41]). They, however, did not explore how emergent leadership interplays with other emergent states. Some computational studies distinctly predict leadership and other emergent states or concepts in order to their correlation (e.g., leadership with dominance [43]), while only a few predict them together (e.g., leadership with cohesion [52]). These studies, however, did not investigate the dynamic aspect of emergent states nor how their approaches could be integrated into other computational models. There are, to the best of our knowledge, no approaches to integrate emergent leadership into other computational models of cohesion.

In this study, we present 2 families of approaches inspired by Social Sciences' insights to integrate emergent leadership into computational models of cohesion over time. Here, our approaches are applied and assessed on a computational model of cohesion since

it is a multidimensional affective emergent state that interplays with leadership and that develops over time [9, 16, 34]. The computational model of cohesion consists of a Deep Neural Network model predicting the dynamics of the Social and Task dimensions of cohesion across various tasks over time.
The first family of approaches (called "*Features Based Leadership*") is based on studies from Psychology showing that the leader is the most influential and active person in the group (e.g., the person who talks and moves the most [3, 13, 19, 47]). For this family, we introduce 2 different approaches that act on the existing features of cohesion. The first one is using a *Normalization* based on leaders' features. The second one is using a *Weighting* strategy on the relevant features of the leader. The second family of approaches (called "*Representation Based Leadership*") focuses on adding leadership representation into the cohesion model architecture. Such information can be treated as an additional feature given to the model. Similarly, 2 different approaches are developed for this family. The first one is adding a leadership representation *Extracted from self- and external assessments* of leadership, while the second one is adding the leadership representation *Automatically Learnt* beforehand using a pre-trained model. Approaches from both families are independently applied to the cohesion model. To assess whether or not integrating leadership into cohesion following our approaches improves cohesion prediction, the performance of each approach is compared against the performance of the same cohesion model that does not apply any approach.

The remainder of this paper is organized as follows. Section 2 describes the related work on leadership in Social Sciences studies and on automated detection of emergent leadership in Social Signal Processing studies. Section 3 gives an overview of the data and the computational model of cohesion used in this study. Section 4 presents the different approaches to integrate leadership into cohesion computational model. Section 5 explains the evaluation methodology of the cohesion model, while Section 6 shows and discusses the results. Conclusions are drawn in Section 7.

2 RELATED WORK

2.1 Background

2.1.1 Leadership. It is a complex phenomenon that has received a lot of attention from scholars in Social Psychology over the last 50 years [14, 23, 35, 48]. This is primarily due to its relationships with team effectiveness and team performance [57]. Leadership is inherently dynamic (i.e., it develops over time) and various styles of leadership exist (e.g., Autocratic vs. Democratic, Transactional vs. Transformational) [18, 40]. Following the recent trends of flattening organizational hierarchies and self-managed teams, the focus of Social Sciences' research switched towards a more informal representation of leadership [38, 56] leading to new types of informal leadership such as emergent leadership, shared leadership, and collective leadership. These kinds of leaderships arise naturally from group interaction, rather than from a higher authority (e.g., a manager) [23]. The way a leader comes to power (formally or informally) does not influence its style.

In this paper, we focus on emergent leadership, an individual emergent state that evolves over time [20]. It is defined as *"the degree to which an individual with no formal status or authority is perceived by one or more team members as exhibiting leaderlike influence"* [23] and has been positively linked to team performance. Previous studies show that a team with an emergent leader can outperform teams with a formally designed leader [15, 46, 48]. The role of an emerged leader is, however, never settled. It depends on the person's abilities, the need of the group, and the team task [45]. Thus, the nature of the task impacts its emergence (i.e., an emergent leader may appear in a team for a particular task but not for another) [48].

2.1.2 Interplay between leadership and other emergent states. Previous studies from Sociology and Psychology reveal that a link between leadership and other emergent states (e.g., cohesion) or group outcomes (e.g., team performance) exists. For example, López-Zafra et al. show a positive correlation between emotional intelligence and leadership [34]. This correlation was better observed in groups with high cohesion. These findings highlight the importance of cohesion in the emergence of a leader through emotional intelligence. In another study, Callow et al. [9] empirically demonstrate that leadership is related to cohesion. They show, indeed, a positive correlation between some of the leadership behaviors (e.g., fostering acceptance of group goals, promoting teamwork) and the Social and the Task dimensions of cohesion. Also, team cohesion has been proven to mediate the relationship between leadership and team performance [16, 49]. In particular, Xie et al. [53] investigated college student teamwork in an online class and showed a strong correlation between emergent leadership and group cohesion. Yamaguchi and Maehr [54] also found that emergent leadership leads to stronger group cohesion in elementary classrooms where students collaborate in math activities.

2.2 Computational Studies

2.2.1 Leadership. Automated detection of emergent leadership is raising interest in the organizational environment as it has been related to the improvement of team productivity and performance [46]. Several datasets were collected to study such an emergent state (e.g., ELEA [43], PAVIS [7]) and used by several studies (e.g., [5, 8, 39]). Since nonverbal features have been found to be more informative than verbal features for the analysis of social phenomena (e.g., [1, 29]) computational studies addressing the automated detection of emergent leadership used nonverbal features (e.g., [6, 43]). Such features can be extracted from different modalities (e.g., from audio, video, or motion capture data). Using the PAVIS dataset, Beyan et al. implemented an automatic method for detecting the visual focus of attention (VFOA) from videos. Based on these, they extracted nonverbal features [7]. They used different types of Support Vector Machines (SVM) (e.g., SVM-cost [25], SVM-SMOTE [11]) as predictive models. Using only the video modality, they achieved, over 3 classes (i.e., the most and the least emergent leader, and the Rest of the group members), 79% accuracy for the most emergent leader, 63% accuracy for the least emergent leader, and 64% accuracy for the Rest class. In a different study [5], the same authors merged the body and head activity features with the visual features and achieved significant improvement for the least emergent leader using SVM (i.e., 72% accuracy). Finally, using a localized multiple kernel learning (LMKL) model, they outperformed previous approaches' performance for the detection of the least emergent leader by 3%. Studies combining both audio and video modalities to extract nonverbal features, however, achieve higher accuracy (e.g., [8, 43, 52]). In particular, Beyan et al. present a meta-analysis of the different nonverbal features used in the emergent leader style detection [6]. They compare the performance of single and multiple modalities features while predicting the existence of an emergent leader and its style (Autocratic vs. Democratic). They tackle this problem as a binary classification task and obtain a score of 0.84 on a geometric mean (GeoMean) metric using LMKL model and multimodal data (i.e., audio and video), which outperforms the models using a single modality [31]. Finally, Beyan et al. integrate the time dimension in their features [8]. Their sequential approach highlights the dynamic aspect of emergent leadership and significantly improves previous results: it obtains an average score over the 3 classes (i.e., the most and the least emergent leader, and the Rest of the group members) of 0.89 on a GeoMean metric. In this study, the authors suggest an approach to generate sequential features from both video and audio using an unsupervised deep learning model. These features are fused afterward and processed by the LMKL model to predict the emergent leader of a group.

As opposed to these different approaches, our work focuses on integrating emergent leadership into a computational model of cohesion instead of focusing on emergent leadership only.

2.2.2 Joint prediction of leadership and other emergent states. To the best of our knowledge, only a few studies simultaneously predict leadership with other group emergent states (e.g., cohesion) or other social concepts (e.g., dominance). In particular, Sanchez-Cortes et al. predict the emergent leader simultaneously with the perceived dominant person [43] since both phenomena are strongly correlated according to Social Psychology [26]. They define the emergent leader as the person who influences other members of the group and contributes to the task solution, whereas the perceived dominant is the person who seeks to stand out and control the others. They also tackle the prediction of the emergent leader and the dominant person as binary classification problems and obtained an accuracy of 85% and 74%, respectively. Additionally, they predict other perceived characteristics (i.e., Perceived Competence and Perceived Liking) that correlate with the emergent leader. Similarly, Zhang et al. predict the emergent leader alongside the major contributor in the group (the person who contributes the most to solve a task) as both are positively correlated to team success [58]. They reach an accuracy of 64% for the binary prediction of emergent leaders and 86% for the binary prediction of major contributors.

To the best of our knowledge, only Wang et al. attempt to integrate cohesion and leadership together by predicting cohesion based on leadership and (dis)agreement between group members [52]. This study, however, only uses audio verbal features with a logistic regression model, which doesn't take into account the dynamic aspects of both phenomena. It also does not compare the performance of the cohesion detection without the leadership influence, making it hard to evaluate the impact of leadership on cohesion.

In our study, we also focus on the integration of leadership into cohesion. Unlike previous studies, we specifically integrate it over time using nonverbal features to explore whether or not integrating such relationships helps to improve the performances of the computational model of cohesion.

3 EXPERIMENTAL SETTINGS

In this paper, we implemented and assessed our approaches on a computational model aimed at predicting the Social and Task cohesion dynamics. Such a model allowed us to investigate the effect of emergent leadership on cohesion as they both develop and interplay over time [4, 16]. We first explain what is the data and the labeling strategy used for emergent leadership detection. Then, we describe the computational model of cohesion that is our baseline against which our approaches are evaluated.

3.1 Dataset

Several multimodal datasets specifically designed for the automated detection of emergent leaders exist (e.g., ELEA [43] and PAVIS [7]). They are, however, composed of groups that only interact once to solve a specific task (e.g., ranking a list of items to survive an airplane crash in winter [43]). These datasets either collected external or self annotations only once. For this reason, they are not able to capture the dynamic aspect of leadership.

As we are interested in integrating the effect of emergent leadership on cohesion over time, we used the GAME-ON dataset [36]. It consists of more than 11 hours of synchronized multimodal recordings (i.e., audio, video, and motion capture data) in which a total of 15 groups of 3 friends interact during an escape game. The game is divided into 5 different tasks, which are listed as follows; Task 1 (T1): searching and finding an object, Task 2 (T2): solving mathematical enigmas, Task 3 (T3): collaborating to solve various clues, Task 4 (T4): guessing the use of an unknown object, and Task 5 (T5): presenting the final solution to escape the room. These tasks were designed to elicit variations of cohesion (i.e., increases or decreases). Since group members considered themselves as friends and no hierarchy exist between them, GAME-ON is suitable for studying emergent leadership.

Additionally, it gathers repetitive self-assessments of both leadership and cohesion and external assessments of the leadership of each group member towards its other group members. These were collected after each task. Leadership annotations were collected through a questionnaire based on [30, 38], using a round-robin rating (i.e., one rate itself and other group members). The leadership questionnaire consists of a set of 5 questions (following Gerpott et al.'s study recommendations [20]) on a 6-points Likert scale ranging from 1 ("*Completely disagree*") to 6 ("*Completely agree*"). Cohesion was evaluated through an adapted Group Environment Questionnaire (GEQ) [10], which assesses the Social and Task dimensions of cohesion, respectively (see [36] for the complete set of questions). GEQ consists of 14 9-point Likert items, ranging from 1 ("*Strongly* disagree") to 9 ("*Strongly* agree") grouped in 4 subscales.

3.2 Labeling strategy for emergent leadership detection

Since more than one person can exhibit leadership in small groups [48], we made the assumption that a group of 3 persons can either be composed of multiple leaders (i.e., 1 or 2) or no leaders. If all the group members were evaluated as leaders, we considered that no clear leaders emerged, hence no leaders exist in the group.

In this study, to determine whether one person is an emergent leader or not, we considered the detection of emergent leadership as a binary classification problem (i.e., 0 means a person is not an emergent leader while 1 indicates it is an emergent leader). Such a strategy helps to differentiate clear leaders from followers. As previously mentioned, group members provided self- and external assessments with a set of 5 questions regarding the leadership questionnaire. Both these assessments have biases [51]. With self-assessment, persons are inclined to judge their performance favorably, while external assessment tends to limit such a bias [27]. Therefore, for each of the 5 questions of the leadership questionnaire, we chose to minimize self-assessments by multiplying them by 0.2 (i.e., the self-rated leadership) and to emphasize external assessments (given by the 2 other group members) by multiplying them by 0.4 (i.e., the external rated leadership). Such values (i.e., 0.2 and 0.4) were empirically chosen so they add up to 1. Then, to compute the leadership scores for each of the 5 questions and for each group member, the self and external rated leadership were summed and normalized by 6 (i.e., the maximum rate possible in the Likert scale). To aggregate individual scores, similar studies that also collected self-assessments, usually averaged the leadership scores of each group member (e.g., [41–43]). Unlike these studies, we chose to use both the mean and the median on the leadership scores previously mentioned. In this way, this labeling strategy captures possible disagreement between group members' ratings, as described in [23]. Afterward, for both the mean and median scores, we applied 2 thresholds to detect the number of emergent leader(s) in each group. A first threshold was applied to identify if at least one leader has emerged. If so, a second threshold distinguished whether 1 or 2 leaders exist in the group. These thresholds were empirically determined. The resulting label distribution was validated by an expert in groups that watched random samples of videos from the GAME-ON dataset to assess leadership. We chose to apply 2 sets of thresholds (i.e., 0.1& 0.05 and 0.09& 0.045 for the first & the second threshold respectively) to the median score as, with these sets, the label's distribution only differs by 8% in edge cases. Using the 2 median scores and the mean score, a majority voting is applied to produce the final label. This labeling strategy resulted in a slightly imbalanced labels distribution (i.e., 60% of leaders and 40% of non-leaders).

3.3 Computational model of cohesion

Automated detection of cohesion is still in its infancy. Studies either used traditional machine learning techniques (e.g., SVM in [24]) to predict cohesion from nonverbal behavior or used deep neural networks focusing on predicting cohesion from images (e.g., [21]). Recently, Maman et al., introduced the *from Individual to Group* (fItG), a deep neural network architecture to predict the dynamics of the Social and Task dimensions of cohesion [37]. They designed their model in such a way that it integrates time dependencies of the features and exploits individual as well as group features. As emergent leadership manifests over time and at an individual level, we chose the fItG as our baseline.
As depicted by Figure 1, the model is composed of 4 modules (i.e., *Features extraction*, *Individual*, *Group* and *Output* modules). We extracted the same features described in [37]. These are multimodal (i.e., extracted from audio and motion capture data) nonverbal features that are computed at both individual and group levels (see

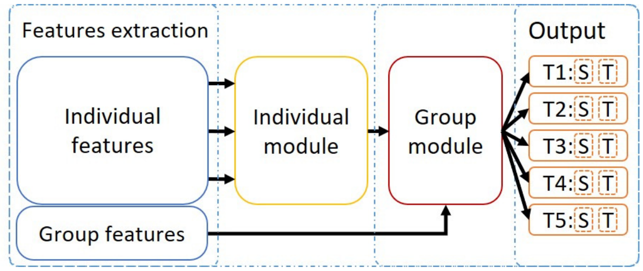

Figure 1: Architecture of the fItG model to predict the dynamics of Social and Task cohesion. It is composed of 4 modules (i.e., Features extraction, Individual, Group, and Output modules). The dynamics of cohesion for the Social (S) and Task (T) dimensions are predicted for the 5 tasks (T1 to T5) in a multilabel setting.

Table 1). The motion capture-based features are composed of features related to proxemics (i.e., the way people use the space) such as the total distance traveled, and related to kinesics (i.e., body movement and gesture) such as the kinetic energy, the posture expansion and the amount of hand gesture. On the other hand, audio features are related to frequency, energy, and turn-taking, which were extracted (see the Geneva Minimalistic Acoustic Parameter Set [17]). The fItG model takes both individual and group multimodal nonverbal features as inputs and it outputs the dynamics of the Social and Task dimensions of cohesion in a multilabel setting, for the 5 consecutive tasks (see Section 3.1). It means that, for each task (T1 to T5 in Figure 1), the model predicts the dynamics of the Social and Task dimensions of cohesion (i.e., decrease vs no decrease) in a multilabel setting

4 PROPOSED APPROACHES

4.1 Features Based Leadership

Scholars in Social Psychology state that a leader is the most influential and active person in the group who talks and moves the most [3, 13, 19, 47]. The 2 following approaches are based on these insights and suggest amplifying the leader's features computed at the individual level. The former, *Normalization*, consists of normalizing the individual features of each group member regarding the ones of the leader(s). The latter one, *Weighting*, gives weight to the relevant leader's features used to predict cohesion.

4.1.1 Normalization. We amplified the differences between the leader(s) and the follower(s) by normalizing each individual's features with respect to the ones of the leader(s). Concretely, we applied the Min-Max scaling method to each individual feature as follows in Equation 1:

$$X_{normalized} = \frac{X - min}{max - min} \tag{1}$$

Where X is the feature vector of a group member, and $X_{normalized}$ is the same feature vector normalized according to min and max. Min and max were chosen based on the feature vector of the emergent leader rather than the min and max of the feature vectors of every group member. When the min and max of the leader's feature vector are not the extremes of every group member, some values of the

Table 1: List of the motion capture-based and audio nonverbal features computed. The features with a '⋆' are computed for the group as a whole while the others are computed for each group member.

Motion capture		Audio	
Proxemics	**Kinesics**	**GeMAPS**	**Turn-taking**
Maximum distance between group members⋆ Total distance traveled Spatial association of max distance⋆ Distance from group Presence of F-Formation⋆	Amount of hand gesture while not walking⋆ Amount of walking⋆ Kinetic energy Synchrony⋆ Posture expansion Touch⋆	Pitch | Jitter F1, F2, F3 frequency F1, F2, F3 relative energy F1 bandwidth Loudness Spectral slope Harmonic differences	Average turn duration⋆ Total speaking time Overlap⋆ Laughter duration

feature vector of the followers are not in the standard range of normalization (i.e., between 0 and 1). For that reason, all the values greater than 1 and less than 0 were set to 1 and 0, respectively. This rounding was made to facilitate the training of the deep learning models.

4.1.2 Weighting. We selected a subset of the features used in the fItG model that are relevant to the emergent leader's behavior (see Table 2). The selection of this subset was inspired by the fact that a leader is perceived by his peers as a dominant person with the most active body language [19]. In more detail, the leader is perceived as the person who walks and talks the most, has an active posture, and is also the person who has the longest variation in the tone of voice and energy [19, 43]. Our approach, here, suggests to only weight this subset of features before inputting it into the model. In that way, the differences between the features of the leader(s) and its follower(s) are amplified. We empirically tested multiple weighting values (i.e., from 1.5 to 5) to observe whether and how amplifying the differences between the emergent leader and its followers improves performances.

4.2 Representation Based Leadership

Unlike the previous family of approaches that acts on existing individual features, this family of approaches aims to integrate a leadership representation into the deep neural network model architecture. Here, we specifically focused on altering the individual module of the fItG since emergent leadership is an individual-level phenomenon [23, 48]. To this aim, we present 2 approaches: the first one, *Extracted from Assessments*), directly uses the leadership scores obtained through the labeling strategy (see Section 3.2) while the second one, called *Automatically Learnt*) uses a representation learned by a pre-trained model that predicts emergent leadership. In both approaches, the leadership representation was concatenated with the output of the individual module from the fItG and became the input of a new Fully Connected layer (FC) shared among the 3 group members. This extra layer allows the model to learn a global higher-level representation of an individual that integrates leadership knowledge. Figure 2 shows how both approaches from this family are integrated into the fItG.

4.2.1 Extracted from Assessments. The first approach re-used the leadership scores computed in Section 3.2. These represent the degree of a person to be perceived as a leader by her peers through the game. They are inserted into the individual module of the fItG model, as shown in the first approach in Figure 2.

4.2.2 Automatically Learnt. Regarding the second approach, the goal is also to add new leadership information by creating an automatically generated leadership representation. To this aim, we designed and trained a deep learning model aimed at predicting emergent leadership. This model was used as a feature extractor and provided the leadership representation of each individual using the features that have been used in automated emergent leadership detection. Using deep learning models as features extractor is a common and robust practice (e.g., [33, 44]). We studied the detection of emergent leadership for each particular individual as a binary classification problem (i.e., 0 means this individual is not an emergent leader while 1 indicates it is an emergent leader). Grounding on [5, 7, 43], we computed 15 features related to the speaking activity (SpeakAct) and 7 features for the Visual Focus Of Attention

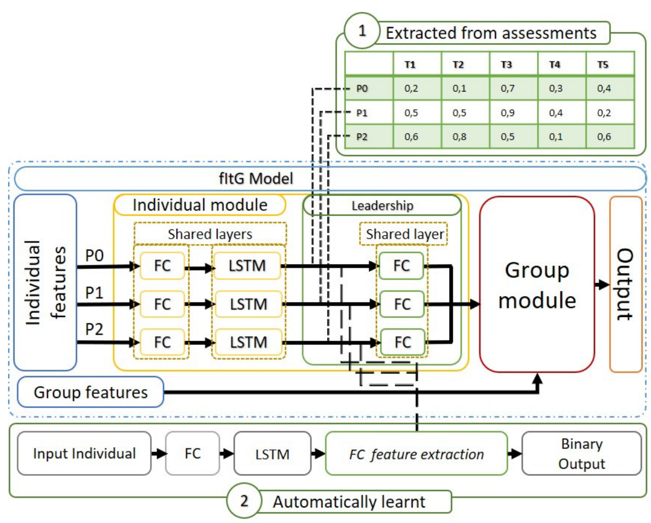

Figure 2: Architecture to integrate the 2 approaches from the Representation Based Leadership family into the fItG model: i) the *Extracted from Assessments* is using leadership scores, ii) the *Automatically Learnt* is generated using a deep learning model. Both of them are integrated into the cohesion model in the individual module using a shared FC layer after the concatenation.

Table 2: Subset of the multimodal features (i.e., motion capture data and audio data) used in the fItG model that are associated with a leader's behavior. Both given names and descriptions are listed (the description for audio modality is taken from [17]). These features are weighted to amplify the differences between emergent leader(s) and follower(s).

Motion capture		Audio	
Feature Name	**Description**	**Feature Name**	**Description**
Total distance traveled	Total distance traveled for a person	Pitch	Logarithmic F0 on a semitone frequency scale
Kinetic energy	Total kinetic energy (translational and rotational)	Jitter	Deviations in individual consecutive F0 period lengths.
Posture expansion	Volume of the space taken by a person	Loudness	Estimate of perceived signal intensity from an auditory spectrum
Amount of hand gesture	Total count of hand gestures while not walking	Total speaking time	Amount of speaking time for a person
Amount of walking	Total amount of time walking	Shimmer	Difference of the peak amplitudes of consecutive F0 periods
		Harmonics-to-noise	Relation of energy in harmonic components to energy in noise-like components

(VFOA) as presented in Table 3. SpeakAct features are related to the speaking length, the interruption between individuals, and the turn-taking of each person, while VFOA features are essentially related to mutual engagement (ME) that is happening when 2 persons are looking at each other at the same time. This set of features was the input of the pre-trained leadership model. This model is composed of a Fully Connected layer with a ReLu activation function and 24 units followed by an LSTM layer and a Fully Connected layer with a ReLu activation function and 16 units. The output of this layer is used to (1) make the final prediction (i.e., leader or not leader) during training thanks to a Fully Connected layer with a sigmoid activation function and 1 unit, and (2) insert the learnt representation of leadership into the fItG model during its training phase. As opposed to the fItG model that processes all individuals at once and predicts both Social and Task cohesion for the 5 consecutive tasks, this pre-trained model that predicts emergent leadership, only processes the features of each person (independently of its group) and predicts the emergence of a leader for a specific task. We trained the model using a 5-fold cross-validation method and a grid-search optimization on both learning rate and epochs to obtain the best combination. The weighted F1-score metric was chosen as it accounts for label imbalance in the data [22] and it allows us to compare the model's performance to the fItG. While running this model using different seeds, we chose the best performing one with an F1-score of 0.72 as the pre-trained model used to extract the leadership representation. Finally, to provide a reliable assessment of this leadership model, we evaluated it using 5-fold cross-validation and a fixed learning rate of 0.0001 coupled with an early stopping regularization technique on the epochs to avoid over-fitting. Then, we averaged its performances over 1000 randomly extracted seeds. The model obtained an averaged F1-score of 0.64 ±0.02. Considering the variety of tasks on which the model is evaluated, such a performance is acceptable.

5 EVALUATION

For each model (i.e., the fItG model without and with our approaches applied) evaluation, we used a Leave-One-Group-Out (LOGO) cross-validation. At each round of the LOGO, a grid-search approach was adopted to select the learning rate (in 0.01, 0.001, 0.0001) and the number of epochs (in 200, 300, 500) on 4 randomly picked groups. To adjust unbalanced labels of cohesion, the loss function was weighted during training inversely proportionally to the classes frequencies, putting more importance on underrepresented classes. To provide a reliable assessment of the models, each model was trained on 15 random seeds [12]. Performances were then averaged. To evaluate if applying our approaches significantly improved the fItG performances, we ran statistical analyses between each family of approaches and the fItG model. A significance level of $\alpha = 0.05$ was chosen. To assess whether or not the differences in performance between models are significant, we used computationally-intensive randomization tests. These are non-parametric tests avoiding the independence assumption between the results being compared and are suitable for non-linear measures such as F1-score [55]. In case of significant difference, a posthoc analysis was carried out using pairwise permutation through False Discovery Rate (FDR) method. In the remainder of this paper, statistical analysis refers to these tests.

6 RESULTS AND DISCUSSION

Results presented in this Section aim to show whether and how applying our approaches improves the fItG performance on both the Social and Task dimensions of cohesion. We first tested if there were significant differences between each family of approaches with respect to the fItG. Then, the best approach from each family were compared between each other.

Regarding the Social dimension of cohesion, there are no significant differences in performance between the approaches from the same family and the fItG. Significant improvements of the F1-Score for

Table 3: List of nonverbal features used in the *Automatically Learnt* leadership representation approach. These features are extracted for each individual separately and are related to their speaking activity (SpeakAct) and their Visual focus of attention (VFOA).

SpeakAct	VFOA
Total speaking length (alone or not)	Total times of looking at anyone without mutual engagement (ME)
Total and Average times of speaking turns	Total times of being looked at by at least one person without ME
Total times of (un)successful interruptions someone	Total times of being looked at without ME
Total times of being (un)successfully interrupted by anyone	Total times of ME with any participant
Total times of speaking first right after anyone	Total times of being looked at with and without ME
Ratio between the total speaking time and silence	total times ones initiate ME with someone else
Ratio between the number of (un)successful interruptions and the total number of speaking turns	Ratio between the number of times being looked at and looking at

the Task dimension of cohesion are, however, achieved for each family of approaches. This result is in line with Social Psychology's insights stating that, when a team is working under a time constraint, emergent leaders focus on the task by assigning roles to the group member and developing strategies to improve team performance [15, 48].
For the sake of clarity, only the significant results are reported, hence, the remaining of the analysis focuses on the Task dimension of cohesion.

6.1 Approaches from the Features Based Leadership family

Concerning the Task dimension of cohesion, statistical tests show a significant difference between this family of approaches and the fItG (p = .010). Post-hoc analysis reveal that only the *Weighting* approach significantly improved fItG's performance (i.e., from 0.61 ±0.03 to 0.64 ±0.04, p = .006). These findings suggest that amplifying a leaders' behavior might be beneficial to a certain extent. In fact, the *Normalization* approach amplifies the differences between the leader's features and its followers' features by 21%. This amplification may give the emergent leader too much importance making followers insignificant to the model. In comparison, the *Weighting* approach (with a weight set to 1.5) amplifies the differences by 4%. We empirically confirmed this effect by using different weights (from 1.5 to 5), which corresponds to an amplification of differences ranging from 4% to 13%. As displayed by Figure 3, amplifying leaders' behavior significantly improves performances (see p-values in the yellow boxes) until a weight of 3 (i.e., an amplification of differences of 11%). Augmenting the weight to a higher value does not improve fItG performances. These results show that highly amplifying the differences between an emergent leader and its followers might go against the emergence of an informal leader. Particularly, a highly amplified leader can be perceived as autocratic (i.e., a leader who has too much control over the task), which has been shown to be less effective during tasks completion [32].

6.2 Approaches from the Representation Based Leadership family

About this family of approaches, statistical tests show a significant difference in performances between both approaches and the fItG for the Task dimension only (p = .001). Post-hoc analysis reveals that both approaches significantly improve fItG performances. In particular, the *Extracted from Assessments* approach based on the computed leadership scores significantly improves fItG from 0.61 ±0.03 to 0.65 ±0.03 (p = .003). The *Automatically Learnt* approach based on the pre-trained model also reaches a significantly better F1-score of 0.67 ±0.04 (p = .003). It also significantly outperforms the *Extracted from Assessments* approach (p = .014). These results show that this family of approaches improves the prediction of Task cohesion. These improvements highlight the benefits of integrating leadership information directly into the computational model architecture. Such approaches help the deep learning model to learn a high-level representation of individuals that integrates leadership's characteristics. Furthermore, the fact that the *Automatically Learnt* approach outperforms the *Extracted from Assessments* indicates that the cohesion model is sensitive to the variety of information added. In particular, the *Automatically Learnt* approach

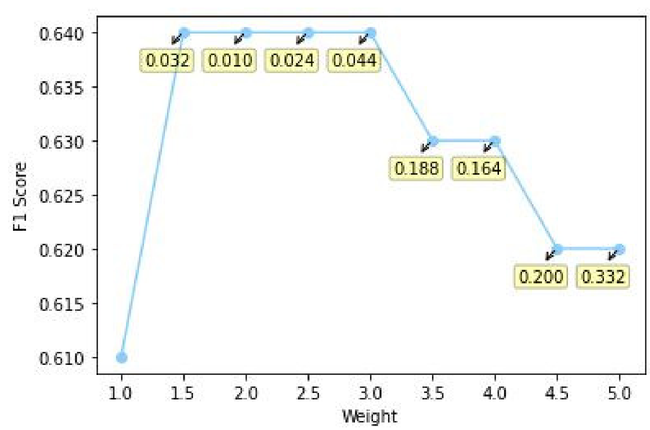

Figure 3: Averaged F1-score over 15 randomly extracted seeds of the fItG model for the prediction of the Task dimension of cohesion with the *Weighting* approach. Weights ranging from 1 to 5 are applied to the features. For each weight, p-values (in yellow) indicate whether or not there is a significant difference with the baseline (weight 1).

Table 4: Summary of the weighted F1-scores for the fItG model and each approach from the Features Based Leadership and Representation Based Leadership families of approaches. The highest F1-scores are in bold.

	Methods	Social F1-score (± std)	Task F1-score (± std)	Avg. F1-score (± std)
	Baseline (fItG)	**0.69 ±0.03**	0.61 ±0.03	0.65 ±0.02
Features Based Leadership	Normalization	0.66 ±0.04	0.62 ±0.04	0.64 ±0.03
	Weighting (by 1.5)	0.68 ±0.03	0.64 ±0.04	0.66 ±0.03
Representation Based Leadership	Extracted from Assessments	0.67 ±0.04	0.65 ±0.03	0.66 ±0.02
	Automatically Learnt	0.67 ±0.03	**0.67 ±0.04**	**0.67 ±0.02**

has a leadership representation for each group member and each task while the *Extracted from Assessments* approach only provides a representation of leadership that is the same across the tasks, for each person.

6.3 Comparing both families of approaches

Lastly, we compare the best-performing approaches of each family, for the Task dimension of cohesion. The *Automatically Learnt* approach based on the integration of leadership representation using a pre-trained model aimed at predicting emergent leadership achieves an F1-score of 0.67 ±0.04. It significantly outperforms the performances of the *Weighting* approach based on applying a weight to the leaders' features that reaches an F1-score of 0.64 ±0.04 (p = .010). This result shows that the *Representation Based Leadership* family of approaches is more effective than the *Features Based Leadership*, highlighting the benefits of adding extra information for learning a representation of individuals instead of solely relying on amplifying existing features. The best performing approach is, indeed, using additional information from other features to automatically detect emergent leaders. Such an approach helps the fItG learning a more complex representation of individuals since it merges 2 sources of information (instead of 1 for the other family of approaches). In that way, the fItG learns new patterns that improve the prediction of the dynamics of the Social and Task dimensions of cohesion. Table 4 summarizes the F1-scores obtained by each approach of both families for the Social and Task dimensions of cohesion as well as the averaged F1-score over both dimensions.

7 CONCLUSION

In this paper, we presented 2 different families of approaches to integrating emergent leadership into computational models of cohesion. The *Features Based Leadership* family consists of amplifying the differences between emergent leaders and followers by acting on the features of the leader. The second family of approaches, named *Representation Based Leadership*, focuses on adding leadership information and representation in deep neural network architectures. We concretely tested both families of approaches on a deep neural network architecture that predicts the dynamics of the Social and Task dimensions of cohesion over time. For each family, we presented 2 different approaches and applied them to the "from Individual to Group" (fItG) cohesion model. Approaches from the same family were evaluated against the fItG's performances. Results show that approaches from both families significantly improve performances for the Task dimension of cohesion. This is in line with Social Psychology's insights regarding the *task-focused* role of an emergent leader stating that a leader assists team performances by assigning roles and developing strategies to complete a specific task [48]. Such a task-focused role might also be emphasized by multiple factors. The fact that, in the studied groups, strong and established relationships already exist between group members (i.e., they consider themselves as friends) could push an emergent leader to focus on accomplishing the group goal. Such behavior is also elicited by the context of the interaction (i.e., during an escape game) that encourages group members to organize themselves to solve different tasks. From a computational point of view, results highlight the benefits of considering the impact of a leader in a group at both feature and model architecture levels. Therefore, integrating such information helps the model learning a higher-level representation of an individual, hence improving the prediction of group cohesion's dynamics.

The work presented in this paper has some limitations. Regarding the *Features Based Leadership* family of approaches, another normalization technique (e.g., z-Score Normalization) could be developed in order to reduce the amplification of the differences between leaders' and followers' features obtained with the current normalization technique. Concerning the *Representation Based Leadership* family, performances of the leadership pre-trained model could be improved by extending the set of features (e.g., the synchronization between the emergent leader and followers as described in [50]) or by using a different architecture (e.g., using a late fusion of the modalities) in order to learn a more robust emergent leadership representation.

ACKNOWLEDGMENTS

This work has been partially supported by the French National Research Agency (ANR) in the framework of its JCJC program (GRACE, project ANR-18-CE33-0003-01, funded under the Artificial Intelligence Plan).

REFERENCES

[1] Dane Archer and Robin M Akert. 1977. Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of personality and social psychology* 35, 6 (1977), 443.
[2] Bruce J Avolio, Fred O Walumbwa, and Todd J Weber. 2009. Leadership: Current theories, research, and future directions. *Annual review of psychology* 60 (2009), 421–449.
[3] John E Baird Jr. 1977. Some nonverbal elements of leadership emergence. *Southern Speech Communication Journal* 42, 4 (1977), 352–361.
[4] Bernard M Bass, Bruce J Avolio, Dong I Jung, and Yair Berson. 2003. Predicting unit performance by assessing transformational and transactional leadership. *Journal of applied psychology* 88, 2 (2003), 207–218.
[5] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2016. Identification of emergent leaders in a meeting scenario using multiple kernel learning. In *Proceedings of the 2nd Workshop on Advancements in Social Signal Processing for Multimodal Interaction*. ACM Digital Library, 3–10.

[6] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2017. Prediction of the leadership style of an emergent leader using audio and visual nonverbal features. *IEEE Transactions on Multimedia* 20, 2 (2017), 441–456.

[7] Cigdem Beyan, Nicolò Carissimi, Francesca Capozzi, Sebastiano Vascon, Matteo Bustreo, Antonio Pierro, Cristina Becchio, and Vittorio Murino. 2016. Detecting emergent leader in a meeting environment using nonverbal visual features only. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM Digital Library, 317–324.

[8] Cigdem Beyan, Vasiliki-Maria Katsageorgiou, and Vittorio Murino. 2019. A sequential data analysis approach to detect emergent leaders in small groups. *IEEE Transactions on Multimedia* 21, 8 (2019), 2107–2116.

[9] Nichola Callow, Matthew J Smith, Lew Hardy, Calum A Arthur, and James Hardy. 2009. Measurement of transformational leadership and its relationship with team cohesion and performance level. *Journal of applied sport psychology* 21, 4 (2009), 395–412.

[10] Albert V Carron, W Neil Widmeyer, and Lawrence R Brawley. 1985. The Development of an Instrument to Assess Cohesion in Sport Teams: The Group Environment Questionnaire. *Journal of Sport Psychology* 7, 3 (1985), 244–266.

[11] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[12] Cédric Colas, Olivier Sigaud, and Pierre-Yves Oudeyer. 2018. How Many Random Seeds? Statistical Power Analysis in Deep Reinforcement Learning Experiments. arXiv:1806.08295 [cs.LG]

[13] Annick Darioly and Marianne Schmid Mast. 2014. The role of nonverbal behavior in leadership: An integrative review. *R. E. Riggio & S. J. Tan (Eds.)* Leader interpersonal and influence skills: The soft skills of leadership (2014), 73–100.

[14] David Day. 2012. Leadership. *The Oxford Handbook of Organizational Psychology* 1 (2012), 696–729.

[15] Gita De Souza and Howard J Klein. 1995. Emergent leadership in the group goal-setting process. *Small group research* 26, 4 (1995), 475–496.

[16] Shelley D Dionne, Francis J Yammarino, Leanne E Atwater, and William D Spangler. 2004. Transformational leadership and team performance. *Journal of organizational change management* 17, 2 (2004), 177–193.

[17] Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. 2015. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE transactions on affective computing* 7, 2 (2015), 190–202.

[18] Franco Gandolfi and Seth Stone. 2017. The emergence of leadership styles: A clarified categorization. *Revista De Management Comparat International* 18, 1 (2017), 18–30.

[19] Fabiola H Gerpott, Nale Lehmann-Willenbrock, Jeroen D Silvis, and Mark Van Vugt. 2018. In the eye of the beholder? An eye-tracking experiment on emergent leadership in team interactions. *The Leadership Quarterly* 29, 4 (2018), 523–532.

[20] Fabiola H Gerpott, Nale Lehmann-Willenbrock, Sven C Voelpel, and Mark Van Vugt. 2019. It's not just what is said, but when it's said: A temporal account of verbal behaviors and emergent leadership in self-managed teams. *Academy of Management Journal* 62, 3 (2019), 717–738.

[21] Shreya Ghosh, Abhinav Dhall, Nicu Sebe, and Tom Gedeon. 2019. Predicting group cohesiveness in images. In *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[22] Cyril Goutte and Eric Gaussier. 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European conference on information retrieval*. Springer Berlin Heidelberg, 345–359.

[23] Andrew A Hanna, Troy A Smith, Bradley L Kirkman, and Ricky W Griffin. 2021. The emergence of emergent leadership: a comprehensive framework and directions for future research. *Journal of Management* 47, 1 (2021), 76–104.

[24] Hayley Hung and Daniel Gatica-Perez. 2010. Estimating Cohesion in Small Groups Using Audio-Visual Nonverbal Behavior. *IEEE Transactions on Multimedia* 12, 6 (2010), 563–575.

[25] Arya Iranmehr, Hamed Masnadi-Shirazi, and Nuno Vasconcelos. 2019. Cost-sensitive support vector machines. *Neurocomputing* 343 (2019), 50–64.

[26] Akko P Kalma, Lieuwe Visser, and Allerd Peeters. 1993. Sociable and aggressive dominance: Personality differences in leadership style? *The Leadership Quarterly* 4, 1 (1993), 45–64.

[27] Linda Koopmans, Claire Bernaards, Vincent Hildebrandt, Stef van Buuren, Allard J Van der Beek, and Henrica CW de Vet. 2013. Development of an individual work performance questionnaire. *International journal of productivity and performance management* 62, 1 (2013), 6–28.

[28] Steve WJ Kozlowski. 2015. Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review* 5, 4 (2015), 270–299.

[29] Uliyana Kubasova, Gabriel Murray, and McKenzie Braley. 2019. Analyzing Verbal and Nonverbal Features for Predicting Group Performance. In *Proceedings Interspeech 2019*. ISCA, 1896–1900.

[30] Klodiana Lanaj and John R Hollenbeck. 2015. Leadership over-emergence in self-managing teams: The role of gender and countervailing biases. *Academy of Management Journal* 58, 5 (2015), 1476–1494.

[31] Jimmie D Lawson and Yongdo Lim. 2001. The geometric mean, matrices, metrics, and more. *The American Mathematical Monthly* 108, 9 (2001), 797–812.

[32] Kurt Lewin, Ronald Lippitt, and Ralph K White. 1939. Patterns of aggressive behavior in experimentally created "social climates". *The Journal of social psychology* 10, 2 (1939), 269–299.

[33] Bing Liu, Xuchu Yu, Pengqiang Zhang, Anzhu Yu, Qiongying Fu, and Xiangpo Wei. 2017. Supervised deep feature extraction for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* 56, 4 (2017), 1909–1921.

[34] Esther López-Zafra, Rocio Garcia-Retamero, and José M Augusto Landa. 2008. The role of transformational leadership, emotional intelligence, and group cohesiveness on leadership emergence. *Journal of Leadership Studies* 2, 3 (2008), 37–49.

[35] Kevin B Lowe and William L Gardner. 2000. Ten years of the leadership quarterly: Contributions and challenges for the future. *The leadership quarterly* 11, 4 (2000), 459–514.

[36] Lucien Maman, Eleonora Ceccaldi, Nale Lehmann-Willenbrock, Laurence Likforman-Sulem, Mohamed Chetouani, Gualtiero Volpe, and Giovanna Varni. 2020. Game-on: A multimodal dataset for cohesion and group analysis. *IEEE Access* 8 (2020), 124185–124203.

[37] Lucien Maman, Mohamed Chetouani, Laurence Likforman-Sulem, and Giovanna Varni. 2021. Using Valence Emotion to Predict Group Cohesion's Dynamics: Top-down and Bottom-up Approaches. In *To appear in Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE.

[38] Elizabeth J McClean, Sean R Martin, Kyle J Emich, and Col Todd Woodruff. 2018. The social consequences of voice: An examination of voice type and gender on status and subsequent leader emergence. *Academy of Management Journal* 61, 5 (2018), 1869–1891.

[39] Philipp Matthias Muller and Andreas Bulling. 2019. Emergent Leadership Detection Across Datasets. In *2019 International Conference on Multimodal Interaction*. ACM Digital Library, 274–278.

[40] Terakanambi Shivashankar Nanjundeswaraswamy and Devappa Renuka Swamy. 2014. Leadership styles. *Advances in management* 7, 2 (2014), 57–62.

[41] Dairazalia Sanchez-Cortes, Oya Aran, and Daniel Gatica-Perez. 2011. An audio visual corpus for emergent leader analysis. In *Workshop on multimodal corpora for machine learning: taking stock and road mapping the future, ICMI-MLMI*. Citeseer, ACM Digital Library, 1–4.

[42] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2010. Identifying emergent leadership in small groups using nonverbal communicative cues. In *International conference on multimodal interfaces and the workshop on machine learning for multimodal interaction*. ACM Digital Library, 1–4.

[43] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2011. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia* 14, 3 (2011), 816–832.

[44] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 815–823.

[45] Anson Seers. 1989. Team-member exchange quality: A new construct for role-making research. *Organizational behavior and human decision processes* 43, 1 (1989), 118–135.

[46] Brian R Spisak, Michael J O'Brien, Nigel Nicholson, and Mark van Vugt. 2015. Niche construction and the evolution of leadership. *Academy of Management Review* 40, 2 (2015), 291–306.

[47] R Timothy Stein and Tamar Heller. 1979. An empirical analysis of the correlations between leadership status and participation rates reported in the literature. *Journal of Personality and Social Psychology* 37, 11 (1979), 1993–2002.

[48] Simon Taggar, Rick Hackew, and Sudhir Saha. 1999. Leadership emergence in autonomous work teams: Antecedents and outcomes. *Personnel Psychology* 52, 4 (1999), 899–926.

[49] Hui-Ling Tung and Yu-Hsuan Chang. 2011. Effects of empowering leadership on performance in management team: Mediating effects of knowledge sharing and team cohesion. *Journal of Chinese Human Resources Management* 2, 1 (2011), 43–60.

[50] Giovanna Varni, Gualtiero Volpe, and Antonio Camurri. 2010. A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. *IEEE Transactions on Multimedia* 12, 6 (2010), 576–590.

[51] Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A Survey of Personality Computing. *IEEE Transactions on Affective Computing* 5, 3 (2014), 273–291.

[52] Wen Wang, Kristin Precoda, Raia Hadsell, Zsolt Kira, Colleen Richey, and Gabriel Jiva. 2012. Detecting leadership and cohesion in spoken interactions. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5105–5108.

[53] Kui Xie, Lauren C Hensley, Victor Law, and Zhiru Sun. 2019. Self-regulation as a function of perceived leadership and cohesion in small group online collaborative

learning. *British Journal of Educational Technology* 50, 1 (2019), 456–468.
[54] Ryoko Yamaguchi and Martin L Maehr. 2004. Children's emergent leadership: the relationships with group characteristics and outcomes. *Small Group Research* 35, 4 (2004), 388–406.
[55] Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. *The 18th International Conference on Computational Linguistics (COLING)* 2 (2000), 947–954.
[56] Stephen J Zaccaro, Roseanne J Foti, and David A Kenny. 1991. Self-monitoring and trait-based variance in leadership: An investigation of leader flexibility across multiple group situations. *Journal of applied psychology* 76, 2 (1991), 308–315.
[57] Stephen J Zaccaro, Andrea L Rittman, and Michelle A Marks. 2001. Team leadership. *The leadership quarterly* 12, 4 (2001), 451–483.
[58] Lingyu Zhang, Mallory Morgan, Indrani Bhattacharya, Michael Foley, Jonas Braasch, Christoph Riedl, Brooke Foucault Welles, and Richard J Radke. 2019. Improved visual focus of attention estimation and prosodic features for analyzing group interactions. In *2019 International Conference on Multimodal Interaction*. ACM Digital Library, 385–394.