

Exploiting the Interplay between Social and Task Dimensions of Cohesion to Predict its Dynamics Leveraging Social Sciences

Lucien Maman, Laurence Likforman-Sulem, Mohamed Chetouani, Giovanna Varni

▶ **To cite this version:**

Lucien Maman, Laurence Likforman-Sulem, Mohamed Chetouani, Giovanna Varni. Exploiting the Interplay between Social and Task Dimensions of Cohesion to Predict its Dynamics Leveraging Social Sciences. 23rd ACM International Conference on Multimodal Interaction, Oct 2021, Montreal, Canada. 10.1145/3462244.3479940 . hal-03409892

HAL Id: hal-03409892
https://hal.science/hal-03409892v1

Submitted on 22 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploiting the Interplay between Social and Task Dimensions of Cohesion to Predict its Dynamics Leveraging Social Sciences

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

Laurence Likforman-Sulem
laurence.likforman@telecom-paris.fr
LTCI, Télécom Paris, Institut polytechnique de Paris
Palaiseau, 91120, France

Mohamed Chetouani
mohamed.chetouani@sorbonne-universite.fr
ISIR, Sorbonne University, CNRS UMR7222
Paris, 75252, France

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

ABSTRACT

Emergent states are behavioral, cognitive and affective processes appearing among the members of a group when they interact together. In the last decade, the development of computational approaches received a growing interest in building Human-Centered systems. Such a development is particularly difficult because some of these states have several dimensions interplaying somehow and somewhere over time. In this paper, we focus on cohesion, its dimensions and their interplay. Several definitions of cohesion exist, it can be simply defined as the tendency of a group to stick together to pursue goals and/or affective needs. This plethora of definitions resulted in many different cohesion dimensions. Social and Task dimensions are the most investigated both in Social Sciences and Computer Science since they both play an important role in a wide range of contexts and groups. To the best of our knowledge, however, no previous work on the prediction of cohesion dynamics focused on how these 2 dimensions interplay. We leverage Social Sciences to address this issue. In particular, we take advantage of the importance of Social cohesion for creating flexible and constructive relationships to reinforce Task cohesion. We describe a Deep Neural Network architecture (DNN) for predicting the dynamics of Task cohesion by applying transfer learning from a pre-trained model dedicated to the prediction of Social cohesion dynamics. Our architecture is evaluated against several baselines. Results show that it significantly improves the predictions of the Task cohesion dynamics, confirming the benefits of integrating Social Sciences insights into models architectures.

CCS CONCEPTS

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

KEYWORDS

Cohesion, Group Dynamics, Multimodal Interaction, Social Signal Processing, Transfer Learning

ICMI '21, October 18–22, 2021, Montréal, QC, Canada
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI '21), October 18–22, 2021, Montréal, QC, Canada*, https://doi.org/10.1145/3462244.3479940.

ACM Reference Format:
Lucien Maman, Laurence Likforman-Sulem, Mohamed Chetouani, and Giovanna Varni. 2021. Exploiting the Interplay between Social and Task Dimensions of Cohesion to Predict its Dynamics Leveraging Social Sciences. In *Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI '21), October 18–22, 2021, Montréal, QC, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3462244.3479940

1 INTRODUCTION

"The whole is greater than the sum of its parts" [2]. This statement by Aristotle, a philosopher in Ancient Greece, is usually employed in Social Sciences to address emergent phenomena in groups such as cohesion, transactive memory system and so on, also called *emergent states*. They are dynamic constructs that characterize properties of the group and that result from the interactions among group members (e.g., [26, 34]). Addressing emergent states computationally is an open challenge of Human-Centered Computing (HCC), a research field aimed at developing computational methods to support and assist human endeavours by studying human interactions through multiple facets [10]. This task is made particularly difficult by the multidimensionality of some of these states because dimensions interplay somehow and somewhere over time. This paper focuses on cohesion, its dimensions and their interplay. This emergent state is one of the most commonly studied in Social Sciences [41] and, more recently, also in Computer Science (e.g., [33]). Scholars in Social Sciences provided multiple definitions of cohesion, identifying several dimensions (generally from 2 to 5, see for example [3, 7, 12, 30]). Despite the scholars' disarray on the number and the nature of these dimensions, the Social and Task dimensions are usually retained in most definitions. Both dimensions, indeed, play a role in a wide range of situations (e.g., a group of friends, a classroom, an emergency team and so on), and are especially relevant for studying task-driven groups (i.e., groups that gather for a purpose). The Social dimension refers to the interpersonal bonds that exist between group members, while the Task dimension corresponds to the group members' shared commitment to the task [43]. According to theoretical models, these dimensions are not orthogonal, meaning that they may all influence each other over time (e.g., [7, 43]). For example, the development of social bonds and friendships within the group (related to the Social dimension) may positively impact Task cohesion. Building on these theoretical models, computational studies naturally started to investigate and

develop methods for the automated analysis of Social and Task cohesion. To date, the focus of computer scientists, indeed, has been either on predicting the intensity level (as Low or High) of cohesion without distinguishing among dimensions (e.g., [22]) or on implementing models predicting the presence and the intensity level of a specific dimension (e.g., [37]). To the best of our knowledge, there is, however, no work on the automatic prediction of cohesion that integrates the interplay between its Social and Task dimensions over time.

In this paper, we take a first step towards bridging this gap by grounding on Social Sciences insights. More specifically, we exploit the role played by the Social dimension for creating flexible and constructive relationships leading to the reinforcement of Task cohesion (e.g. [43], [47]). Concretely, we present a Deep Neural Network (DNN) architecture for predicting the dynamics of cohesion by applying transfer learning. This is done to take advantage of the information learnt by a model dedicated to the classification of Social cohesion's dynamics to predict the Task cohesion's dynamics [46]. Dynamics here refers to changes in cohesion (i.e., decrease or not-decrease). This architecture also takes into account temporality by integrating, amongst others, LSTM layers and it models cohesion at both individual and group levels. The DNN's performances are evaluated against baselines predicting one dimension at a time or predicting both dimensions using traditional machine learning techniques such as multilabel classification.

2 BACKGROUND AND RELATED WORK

2.1 Background

In the 40s, Lewin first defined cohesion as *"a group characteristic that depends on its size, organization and intimacy"* [27]. This definition grounds on the force field theory that views people's activity as affected by forces in their surroundings and environment [28]. Building upon this work, later, scholars provided several definitions of cohesion (e.g., [3, 4, 12, 30]) and multidimensional models of it, in which the number of dimensions varies from 2 to 5. For example, in [6], Carron and Brawley defined the *Group Integration* and the *Individual Attraction To Group* dimensions to take into account the contributions of both each group member and the group as a whole. These 2 dimensions have Social and Task as their sub-dimensions. According to Griffith, depending on the presence (or absence) of hierarchical relationships among the group members, cohesion can be also studied at the horizontal level (e.g., a group of friends) or at the vertical level (e.g., a teacher with her students) [17]. He differentiated 2 dimensions: the *Instrumental* (or Task) and the *Affective* (or Social) dimensions. Bollen and Hoyle integrated other dimensions related to the sense of belonging and the feeling of morale associated with membership in the group [5]. Social and Task dimensions always appear in all this plethora of definitions. Furthermore, Salas *et al.* conducted a meta-analysis in which they recommend giving priority to Social and Task cohesion and integrating time when studying cohesion [42]. Recently, Severt and Estrada proposed a framework of cohesion that gathers all of these efforts to categorize the structural and functional properties of cohesion [43]. This framework posits that cohesion serves an *Affective* and an *Instrumental* function. The former refers to all the aspects that highlight the emotional impact on a group member and, by extension, the group as a whole (e.g., behaviors or elements of interaction such as cooperation or exchange) and is structured into the *Interpersonal* and the *Group Pride* dimensions. The instrumental function corresponds to the aspects that highlight the goal- and task-based activities of the group and is composed of the *Social* and *Task* dimensions. Finally, for each dimension, we can distinguish 2 levels (i.e., horizontal and vertical). This distinction is particularly important as, depending on the dimensions and the level at which cohesion is investigated, it might emerge and evolve differently.

In this study, we follow Severt and Estrada's framework by specifically focusing on the interplay between the Social and Task dimensions, at horizontal level. This level of investigation was chosen as it aligns with the contemporary trends of flattening organizational hierarchies and self-managed teams [31], improving the applicability of our findings.

2.2 Related work

To date, computational studies about cohesion focus on analysing its Social and Task dimensions since they are the easiest to grasp and measure [42]. Hung and Gatica-Perez [22] studied the role played by audio, visual and audio-visual features on the prediction of cohesion intensity level (Low or High) by binary classifiers. Their models addressed cohesion as a whole, that is without distinguishing between its dimensions. Nanninga and colleagues extended this work, integrating pairwise and group features related to the alignment of para-linguistic speech behavior and addressing Social and Task dimensions separately [37]. They found that their audio features such as synchrony and convergence are more relevant to predict the Social dimension rather than the Task dimension. Both these studies show the importance of turn-taking and mimicry features for the prediction of cohesion intensity level. Despite cohesion being inherently temporal, these studies unitized interaction in non-overlapped temporal units which feed models as independent samples. This approach cannot catch dynamics that according to [42] is a relevant point to understand cohesion. Later on, researchers started to investigate Social and Task dimensions of cohesion at a longitudinal level with the use of sociometric badges (i.e., objects placed on a person or on its phone, that are able to track a person's movement and activity) [51]. These badges were used to quantify dyadic interactions and face-to-face communications and analyze small group collaborations during long-duration missions in confined spaces. Recently, a growing interest in the interplay between the several dimensions of cohesion emerged also in Computer Science. To address this issue, a couple of studies developed methods inspired by game theories. In [39], the authors investigated the differences between the Social and Task dimensions using an approach based on evolutionary game theory by promoting the evolution of cooperation in group interactions. They reveal that Social cohesion is detrimental to the evolution of cooperation while Task cohesion facilitates it. They explain these results by exploring the effects of the mistake rate on the cooperation of the groups and show that increasing Task cohesion would preserve cooperation from mistakes while increasing Social cohesion would augment the mistake rate. Their model, however, uses randomly generated data and does not consider the reciprocal effect of both dimensions.

Walocha *et al.*, proposed a method based on notions from cooperative game theory (i.e., using SHAP values) to assess the importance of motion capture-based features on a random forest model predicting the dynamics of the Social and Task dimensions of cohesion. Interplay is faced through a multilabel classification [49]. Their model, however, addresses cohesion using a single modality only.

3 A SOCIAL SCIENCES INSPIRED MODEL FOR INTEGRATING SOCIAL AND TASK INTERPLAY

In this Section, we present the *Transfer Between Dimensions* (TBD) architecture. This is a DNN architecture that: (1) integrates the interplay between Social and Task dimensions of cohesion following Social Sciences insights; (2) takes into account temporality by integrating, amongst others, LSTM layers; and (3) models cohesion at both individual and group levels. In the following, for each of these items, we describe the Social Sciences insights we ground on and how such knowledge is reflected in the architecture. Concerning (1), Carron and Brawley question whether and which one of the Social and Task dimensions is predominant over the other one. They argue that Social cohesion might particularly impact Task cohesion depending on many factors including the context of the interaction, the type of group (e.g., work team) and the stage of formation of the group (e.g., early-stage) [6]. In [47], the authors claim that Social cohesion would likely be more salient in social groups such as group of friends. Also, as suggested by Grossman *et al.* [19], Social cohesion emerges first in the group before its members shift attention to Task cohesion. Furthermore, Severt and Estrada [43] state that Social cohesion may create flexible and constructive relationships and would by extension, reinforce Task cohesion. Hence, Social cohesion is expected to be more salient and to be a driver for Task cohesion [19, 47]. For these reasons, we decided to implement TBD, a transfer learning approach to predict the dynamics of Task cohesion based on a pre-trained model dedicated to the prediction of Social cohesion's dynamics. In that way, the model takes advantage of the Social representation of cohesion previously learnt to optimize Task cohesion prediction. About (2), as cohesion is an emergent state, it is by definition a temporal construct changing over time [18]. This implies that the relationship between the 2 dimensions develops over time. TBD takes into account the temporality into its architecture and also predicts the dynamics of cohesion at multiple points in time. Finally, concerning (3), since cohesion is a group-level phenomenon our architecture needs to model both individual and group behaviours simultaneously. According to Cattell [8], groups can be studied at 3 different levels: individual, structural (interactions within the group) and syntality (group as a whole), highlighting the need to consider individual and group contributions. In TBD, this is done both at the feature level (with a distinction between individual and group features) and at the architectural level (with an individual component and a group component, that are aimed at learning the temporal dynamics of cohesion at individual and group levels, respectively). In the following, we describe a TBD instance working on a data set explicitly conceived to study cohesion over time and its dimensions.

3.1 Data set

Several data sets of social interactions in groups exist (e.g., Panoptic [23], MUMBAI [13], AMI [35]). These data sets were either collected to study social interactions in a specific context (e.g., meeting, board game) or to improve group detection or tracking algorithms. Some of these data sets also offer assessments of specific phenomena such as emotion and leadership provided by the participants or by pools of external observers. These assessments, however, are generally made over time windows defined according to technical constraints or Social Sciences theories, and without a particular focus on the development of the measured phenomenon over time. In this study, we adopt the GAME-ON data set [32]. To the best of our knowledge, this is the only publicly available data set specifically designed for the study of Social and Task cohesion and it provides repeated self-assessments of it over time for each member of the group. A slightly modified version of the Group Environment Questionnaire (GEQ) [7], indeed, was administered between each pair of tasks. GEQ is a well-established questionnaire already used by several studies to measure the Social and Task dimensions separately. Moreover, this data set was conceived following the theoretical findings of scholars in Social Sciences such as [47] and [43]. GAME-ON is a multimodal data set (audio, video, and motion capture recordings) in which small groups of 3 friends interact in the context of an escape game. The data set includes more than 11 hours of interaction involving 15 groups. The average duration of a session is 35min and 30s (SD = 4min 10s). The escape game scenario is structured in 5 tasks, explicitly designed to elicit changes in the Social and Task dimensions of cohesion (i.e., increase or decrease of cohesion with respect to the previous task). The first task lasts about 10 minutes, while the second task lasts about 9 minutes. The third and fourth tasks take 7 minutes and the fifth task 8 minutes. Table 1 describes the expected variations of cohesion between 2 consecutive tasks (T) for both the Social and Task dimensions. In [32], authors showed that, except for the T1-T2 and T2-T3 transitions of the Task dimension, the expected changes in cohesion were confirmed by participants' answers.

3.2 Individual and group features

We developed and extracted a set of 84 motion capture-based and audio nonverbal features characterizing social interaction. For the sake of brevity and narrative clarity, the details of implementation are not given. Features computed from motion capture data concern proxemics (i.e., the way people use the space) and kinesics (i.e.,

Table 1: Expected changes in cohesion between 2 consecutive tasks (T) for the Social and Task dimensions as reported in [32]. Start means the beginning of the recordings.

	Change in cohesion	
Transition	**Social**	**Task**
Start - T1	Decrease (↘)	↘
T1 - T2	↘	Increase (↗)
T2 - T3	↗	↘
T3 - T4	↗	↗
T4 - T5	↗	↗

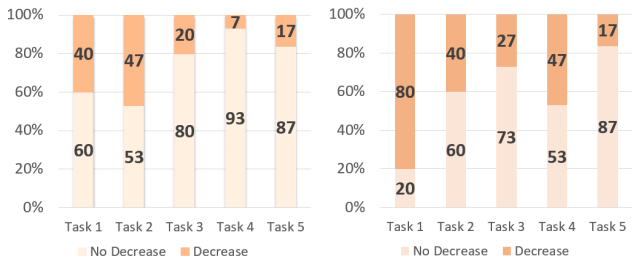
Table 2: List of the motion capture-based and audio nonverbal features characterizing social interaction used in this study. The features with a "⋆" are the ones for which we applied statistical functions (i.e., mean, std, min, max and skewness).

		Individual	Group
Motion Capture	**Proxemics**	Distance from individual to the barycentre of the group ⋆ Total distance traveled	Maximum distance between group members ⋆ Time in F-Formation (triangle or semi circular) Interpersonnal distances (Public, Social and Personnal spaces)
	Kinesics	Posture expansion ⋆ Kinetic energy ⋆	Amount of walking ⋆ Amount of hand gesture while not walking ⋆ Touch detection ⋆ Synchrony of kinetic energies
Audio	**Turn-taking**	Laughter duration Total speaking time	Average turn duration
	GeMAPS	Pitch / Jitter / Loudness Spectral slope / Harmonic differences F1,F2,F3 frequency and relative energy F1 bandwidth	Time of overlapping speech

body movement and gesture). Indeed, both play an important role in nonverbal communication [21] and is relevant for predicting Social and Task dynamics of cohesion [49]. Regarding the audio data, we adopt features from the Geneva Minimalistic Acoustic Parameter Set (GeMAPS; see [15] for a detailed description) that is composed of features related to frequency, energy, spectral balance and temporal features of the voice. These features have been successfully used in many affect related prediction tasks such as predicting cohesion or emotions (e.g., [40]). Furthermore, we included turn-taking related features (e.g., average turn duration), taking inspiration from previous work showing their relevance for the automated detection of cohesion (e.g., [22]). Among all the features used in this study, some are computed for each group member (e.g., the total distance travelled) while others are computed at a group level (e.g., the time of overlapping speech). Features were computed over non-overlapped time windows of 20s according to previous work on group interaction [16] and cohesion perception [9] exploiting a *thin slices* approach. This refers to the process of making very quick inferences about the individual and/or group phenomena with a minimal amount of information [1]. Table 2 summarizes the features we used. A "⋆" indicates that we applied statistical functions (i.e., mean, std, min, max and skewness) instead of using the computed value of the features. For sake of clarity, some of the names chosen for describing the features concern an ensemble of features related to the same behavior (e.g., posture expansion regroups the latitudinal and longitudinal expansion features).

3.3 Labels

Based on the self-assessments of cohesion provided with the GAME-ON data set, we build labels of Social and Task cohesion for each pair of consecutive tasks. We defined a labeling strategy that formalizes the prediction of the dynamics of cohesion as a binary classification problem (decrease vs not-decrease). In this work, we explicitly focused on decreases in cohesion. This, indeed, is an established method in research on Affective Computing and Social Signal Processing (e.g., [36, 45]). Specifically, we rank the 6 scores



(a) Social cohesion (b) Task cohesion

Figure 1: Distribution of the labels (Decrease vs No Decrease), for each task and for the Social and Task dimensions of cohesion (see Figure 1(a) and Figure 1(b), respectively).

provided by the group members (i.e., 2 scores each) and then compute their mean difference. Finally, we binarize the labels based on their sign: a negative label indicates a decrease of cohesion, resulting in assigning a value equal to 0 to the label; when a label is positive, a value of 1 is given (i.e., no decrease of cohesion is observed). Overall, this strategy led to an imbalanced distribution of the labels for the Social dimension (i.e., 73% of "not-decrease" labels) and a balanced distribution for the Task dimension (i.e., 56% of "not-decrease" labels). As displayed by Figure 1, the distribution per task is also highly imbalanced for both dimensions.

3.4 Overall architecture

The TBD architecture consists of 4 modules (see Figure 2). As mentioned above, this model was conceived to take advantage of a pre-trained model dedicated to the classification of Social cohesion's dynamics to predict the Task cohesion's dynamics, using transfer learning. Hence, TBD only focuses on predicting the dynamics for the Task dimension. The dynamics for the Social dimension are predicted during the training phase of the pre-trained model.

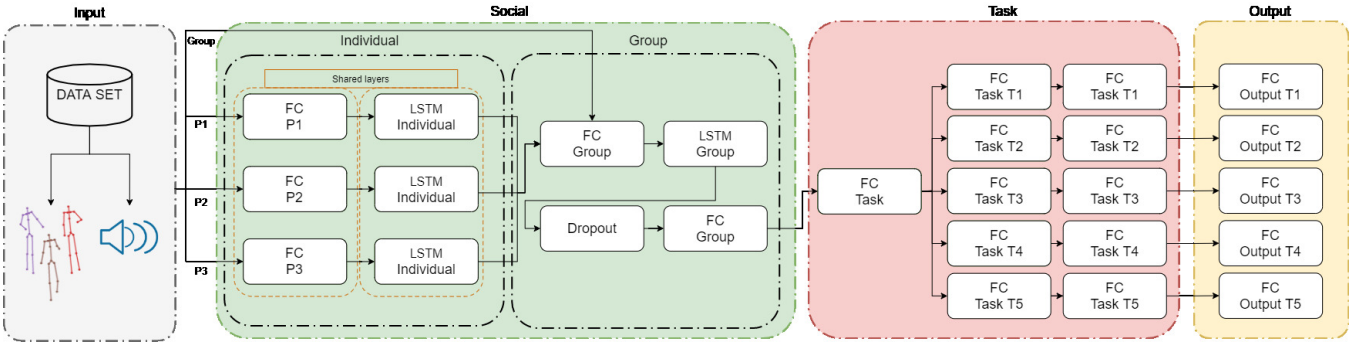


Figure 2: The architecture of the TBD model. It is composed of 4 modules (i.e., *Input, Social cohesion, Task cohesion* and *Output*). Using multimodal features, the model integrates both individual and group contributions. The dynamics of Task cohesion are predicted for each of the 5 tasks, using insights from Social cohesion learnt beforehand. The architecture is composed of fully connected (FC), LSTM and Dropout layers.

3.4.1 Input. This module extracts the individual and group features (see 3.2). In this study, features were computed on non-overlapped time windows of 20s spanning the 2 last minutes of interaction for each of the 5 tasks. This results in a total of 30 windows per group. The choice to focus on the last 2 minutes was motivated by the use of the self-assessments provided by the group members. As reported in several studies carried out in very different contexts, self-assessments collected through questionnaires are likely influenced by the last recalled behavior (e.g., [14, 29]). Then, due to the relatively small size of GAME-ON, the module performs data augmentation to create synthetic groups by permuting the order of the group members of the 15 groups. In this way, the size of the data set is 6 times bigger than initially. Data augmentation also allows us to prevent the model from learning undesirable patterns related to the order in which the model is processing group members.

3.4.2 Social cohesion. This module consists of a pre-trained model to predict Social cohesion. In that way, we use the higher representation of the Social dimension learnt beforehand to start with a better initialisation point. Moreover, as we are interested in the interplay of Social and Task cohesion, the weights of the pre-trained model are modified during the training phase of the TBD so that it also integrates the impact of Task cohesion on Social cohesion. One of the advantages of this pre-trained model is that it uses both individual and group features to learn a higher common representation merging individual as well as group representations to predict Social cohesion. Furthermore, its structure enables the integration of the temporality between the time windows as it is composed of LSTM layers both at individual and group levels. The pre-trained model, also used in [33], consists of 2 components: the *Individual* and the *Group* components. The former has 3 branches (1 per group member), where each one of them is composed of a fully connected (FC) layer with 50 units and a ReLu activation function, followed by an LSTM layer with 50 units. To let the model learn a global representation of an individual, each layer of the 3 individual branches (i.e., the FC and LSTM layers) is shared following Equation 1:

$$\mathbf{Y}_i = \phi\left(\sum_{j=1}^{n}(W\mathbf{X}_j)\right) \tag{1}$$

where Y_i is the output of layer i (i.e., the FC and LSTM layers), ϕ_i, the activation function of the layer i, W, the matrix of parameters common to every group members and X_j, the input related to group member j. As groups are composed of 3 persons, n was here set equal to 3. The Group component is aimed at learning the temporal dynamics of cohesion at the group level. It takes multiple inputs by concatenating the group features with the outputs of the 3 individual LSTM layers from the Individual component. The Group component is made of a first FC layer with 64 units and a ReLu activation function, followed by an LSTM layer with 32 units to integrate the group temporality. Next, a Dropout layer with a rate of 0.2 is used to prevent the model from overfitting. This layer is followed by another FC layer with 16 units and a ReLu activation function.

3.4.3 Task cohesion. The output of the Social cohesion module is used as input of an FC layer with 16 units and a ReLu activation function. In that way, the model learns a higher representation of Task cohesion before splitting into 5 branches (1 for each task) so that each branch learns the task specificity. Branches are composed of 2 FC layers with 8 and 4 units, respectively and a ReLu activation function.

3.4.4 Output. Finally, this module consists of an FC layer with 1 unit and a sigmoid activation function, for each task. The resulting outputs are the predictions of the dynamics of Task cohesion.

4 MODEL EVALUATION

4.1 Method

TBD is evaluated through the following procedure. A 3-fold nested Leave-One-Group-Out (LOGO) cross-validation was carried out. The hyper-parameters, here the learning rate and the number of epoch, were in {0.01, 0.001, 0.0001} and {100, 200, 300, 500}, respectively. The imbalance in the data was handled by automatically weighting the loss function during training, in an inversely proportional way to the class frequencies according to Equation 2. This method is inspired by [24].

$$cw_i = \frac{n_g}{n_c * n_i} \tag{2}$$

where cw_i is the class weight used in the loss function during training for the class i; n_g, the number of groups; n_c, the total number of classes (i.e., *decrease* and *not-decrease*) and n_i, the number of occurrences for the class i.
The model's performances are evaluated using the following 2 metrics: (1) the F1-score per task (i.e., across the 15 rounds of the LOGO) and (2) the mean of the F1-score obtained for the 5 tasks, for each dimension. These metrics account for the label imbalance (e.g., [20]) and give us insights into the ability of the model to correctly predict under-represented classes.

Finally, according to Colas *et al.*'s guidelines [11] that suggest using a number of seeds ranging from 5 to 25 depending on the data and the algorithms, we train our models on 15 different randomly extracted seeds and average the performances to obtain a more robust measure of the architecture performances. In this way, we aim at providing a reliable assessment of the models' performances. In this study, we first compare 3 baselines between each other to select the most performing one and then we compare it to the TBD model. Statistical significant differences between the performances of the models were assessed via computationally-intensive randomization tests using $\alpha = 0.05$. These are non-parametric tests avoiding the independence assumption between the results being compared and that are suitable for non-linear measures such as F1-score [50].

4.2 Baselines

Three different models were used as baselines to predict the dynamics of cohesion for the Social and Task dimensions. Such baselines range from a simple but consolidated state-of-the-art approach to more sophisticated approaches that increasingly address temporality and group contributions. Each baseline has been implemented in 2 different ways: predicting the dynamics of the 2 dimensions of cohesion separately or using a multilabel classification as a first attempt to integrate the interplay between Social and Task cohesion. The latter implies that both dimensions are tightly related to each other since the overall loss is the unweighted sum of the losses from both dimensions and only a final FC layer is differentiating both dimensions, for each task. For the remainder of the paper, the versions of the baselines predicting each dimension separately are mentioned with the "_SD" suffix (as for Single Dimension) while the multilabel versions are with the "_MD" suffix (as for Multiple Dimensions). Each model was evaluated following the procedure described in 4.1, that is by using 3-fold nested LOGO cross-validation with hyperparameters tuning.

4.2.1 Tree based approach. As stated by [48], Random Forest is one of the most powerful algorithms for solving binary classification problems. For this reason, we decided to use this classifier (RFC) as our first baseline to predict the dynamics of cohesion by predicting each of the 20-second thin slices. Since we are using the last 2 minutes of interaction, it means that the RFC makes 6 predictions per task and dimension. A majority voting is then applied over these 6 predictions to determine the overall prediction of the task, for each dimension. At each round of the LOGO cross-validation, a feature selection algorithm based on Kolmogorov-Smirnov statistic [25, 44] is applied to reduce the feature set, as referred to in [38]. The estimated hyperparameters were: the number of trees (in {100, 200, 300, 400, 500}), the maximum depth of the tree (in {10, 20, 30, 40, 50, 60, 70, 80, 90, 100}), the minimum number of samples required to split an internal node (in {1, 2, 3, 4, 5}), and the minimum number of samples required to be at a leaf node (in {2, 3, 4, 5, 6, 7}). RFC, however, does not model the time dependencies between the thin slices nor between the tasks and does not model the group at multiple levels (e.g., individual and syntality levels).

4.2.2 Integrating the time dependencies between the thin slices and between the tasks. To model the time dependencies between the thin slices and between the tasks, we designed the *Full-Interaction LSTM* (FI-LSTM) model. This DNN architecture integrates the temporality by inputting the features to an LSTM layer with 30 units (i.e., the number of thin slices composing the whole interaction). This layer is followed by a Dropout layer with a dropout rate of 0.2 and by 2 FC layers with 16 and 8 units, respectively, and a ReLu activation function. FI-LSTM predicts the dynamics of Social and/or Task cohesion for each of the 5 tasks of an interaction thanks to a final FC layer with 1 unit (or 2 units if the model is using a multilabel classification) and a sigmoid activation function for each task.
As for the RFC, FI-LSTM still does not integrate into its architecture how a group, as well as individuals, contribute to cohesion.

4.2.3 Integrating time dependencies and both individuals and group contributions. The last baseline is the *from Individual to Group* (fItG) model. Similarly to FI-LSTM, this model predicts the Social and Task cohesion's dynamics while integrating the time dependencies between the thin slices and between the tasks. In addition, it learns individual as well as group representation of the features. FItG is used as the pre-trained model for predicting Social cohesion in the TBD architecture. Its architecture is described in Section 3.4.2. The only difference resides in the output of the fItG. Indeed, as for the FI-LSTM, we added 5 distinct FC layers with 1 unit (or 2 units if the model is using a multilabel classification) and a sigmoid activation function (1 for each task).

5 RESULTS AND DISCUSSION

TBD and each baseline were developed and trained using Python 3.7 and Tensorflow 2.1 on NVIDIA V100 GPUs. Table 3 summarizes the performances of the 3 baseline algorithms both when they predict the 2 dimensions of cohesion separately and when they predict them using a multilabel approach. The results are reported per task and per dimension. First, we tested whether the initial attempt of using multilabel classification to take into account the interplay of Social and Task cohesion outperforms the performances obtained by predicting them separately. Then, the best baseline is retained to be tested against TBD.

5.1 Comparing the baselines

5.1.1 Single-label vs Multilabel. RFC_MD significantly improves the prediction of the Social dimension (from 0.61 ±0.01 to 0.62 ±0.02, $p = .044$), while it significantly decreases that one of the Task dimension (from 0.55 ±0.02 to 0.53 ±0.02, $p = .002$). Concerning the 2 other DNN architectures (i.e., FI-LSTM and fItG), no significant difference is found for the Social dimension. Multilabel classification, however, significantly improves the prediction of the Task dimension: FI-LSTM_MD achieves 0.63 ±0.02 vs 0.59 ±0.04 obtained

Table 3: F1-scores obtained for all the baselines predicting a single dimension vs using multilabel classification. Tables on the left and on the right report the performances for the Social and Task dimensions, respectively. Results are displayed per task.

	F1 scores ±std - Social dimension					
	RFC_SD	RFC_MD	FI-LSTM_SD	FI-LSTM_MD	fItG_SD	fItG_MD
T1	0.45 ±0.00	0.47 ±0.06	0.54 ±0.12	0.52 ±0.08	0.53 ±0.09	0.52 ±0.08
T2	0.20 ±0.04	0.23 ±0.04	0.40 ±0.15	0.52 ±0.11	0.53 ±0.11	0.59 ±0.12
T3	0.70 ±0.02	0.70 ±0.02	0.65 ±0.07	0.64 ±0.07	0.60 ±0.07	0.61 ±0.06
T4	0.89 ±0.02	0.87 ±0.00	0.86 ±0.04	0.85 ±0.04	0.87 ±0.03	0.88 ±0.03
T5	0.81 ±0.00	0.83 ±0.05	0.79 ±0.05	0.78 ±0.03	0.80 ±0.04	0.84 ±0.05
Avg	**0.61 ±0.01**	**0.62 ±0.02**	**0.65 ± 0.03**	**0.66 ±0.03**	**0.67 ±0.03**	**0.69 ±0.03**

	F1 scores ±std - Task dimension					
	RFC_SD	RFC_MD	FI-LSTM_SD	FI-LSTM_MD	fItG_SD	fItG_MD
	0.51 ±0.04	0.42 ±0.06	0.54 ±0.08	0.64 ± 0.06	0.54 ±0.11	0.69 ±0.06
	0.37 ±0.05	0.35 ±0.03	0.61 ±0.11	0.66 ±0.08	0.52 ±0.08	0.55 ±0.11
	0.54 ±0.01	0.54 ±0.03	0.59 ±0.06	0.62 ±0.05	0.59 ±0.07	0.60 ±0.09
	0.60 ±0.02	0.61 ±0.02	0.40 ±0.11	0.41 ±0.10	0.43 ±0.13	0.43 ±0.08
	0.76 ±0.04	0.73 ±0.00	0.80 ±0.01	0.80 ±0.01	0.79 ± 0.02	0.78 ±0.03
	0.55 ±0.02	**0.53 ±0.02**	**0.59 ±0.04**	**0.63 ±0.02**	**0.57 ±0.03**	**0.61±0.03**

by FI-LSTM_SD (p = .006), fItG_MD achieves 0.61 ±0.03 vs from 0.57 ±0.03 obtained by fItG_SD (p = .004).

These results show that a simple approach to integrate the interplay of the Social and Task dimensions (i.e., using multilabel classification) partially improves the performances of the models predicting a single dimension. In particular, improvements mainly concern Task cohesion. This shows that a multilabel approach has the potential of improving prediction. Such a kind of approach, however, neglects the insights from the extensive research in Social Sciences that we expect to be beneficial for the model as it implies that both dimensions are strongly related to each other (i.e., they are both predicted from the same node or layer) and equally contribute to cohesion (e.g., by summing the losses of both dimensions for the DNN architectures).

5.1.2 Selecting the best baseline. Since the models using multilabel classification partially improve the performances of the ones predicting a single dimension, we run an extensive analysis on their performances to select the best baseline among the RFC_MD, the FI-LSTM_MD and the fItG_MD. We first compare the performances of the Social and Task dimensions for each model to analyse whether a dimension is easier to predict or not. Then, we compare each dimension, separately, across all the baselines. For each model, the dynamics of Social cohesion are significantly better predicted than the ones of Task cohesion. RFC_MD reaches 0.62 ±0.02 for the Social dimension while it achieves 0.53 ±0.02 for the Task dimension. This difference in the performances is significant (p = .002). FI-LSTM_MD also obtains significantly better results for the Social dimension than for the Task dimension (p = .004) with a F1-score of 0.66 ±0.03 and 0.63 ±0.02, respectively. Similarly, the fItG_MD achieves a significantly better F1-score (p = .002) of 0.69 ±0.03 for the Social dimension with respect to the Task dimension that reaches a F1-score of 0.61 ±0.03.

For every model, the performances of the first 2 tasks are particularly mispredicted. This could be explained by the fact that Social cohesion develops over time and might not manifest during the early stage of the interaction [43]. Concerning the Task dimension, the DNN architectures (i.e., the FI-LSTM_MD and the fItG_MD) particularly mispredicted Task 4. This result could be explained by the nature of this task in which group members had to agree on a solution to solve a quiz. In case of disagreements, group members might have provided very different cohesion scores for the Task dimension, resulting in opposite labels (i.e., decrease vs no decrease) within the same group. This is a limitation of our labeling strategy that does not integrate the potential disagreements within the group, making it harder for the model to predict the dynamics of cohesion for this particular task.

The RFC_MD model achieves, on average over the 15 seeds, a F1-score of 0.62 ±0.02 for the Social dimension and 0.53 ±0.02 for the Task dimension. Statistical analysis shows that there are significant differences in performances for the RFC_MD with respect to the FI-LSTM_MD and the fItG_MD models for both Social (p = .001) and Task (p = .001) dimensions (see also Figure 3). A post-hoc analysis using pairwise permutation t-tests was carried out and showed that both the FI-LSTM_MD and the fItG_MD models outperformed the RFC_MD for the Social (p = .003) and Task (p = .003) dimensions. Indeed, for the Social dimension, the FI-LSTM_MD model reached, on average over the 15 seeds, a F1-score of 0.66 ±0.03 while the fItG_MD obtained an averaged F1-score of 0.69 ±0.03. This improvement of 0.03 observed for the fItG_MD performances is also significant (p = .016). Regarding the Task dimension, the FI-LSTM_MD achieved an averaged F1-score of 0.63 ±0.02 and the fItG_MD reached 0.61 ±0.03. This difference is, however, not significant. To summarize, the fItG_MD model is the most performing one with a F1-score of 0.69 ±0.03 and 0.61 ±0.03 for the Social and Task dimensions, respectively.

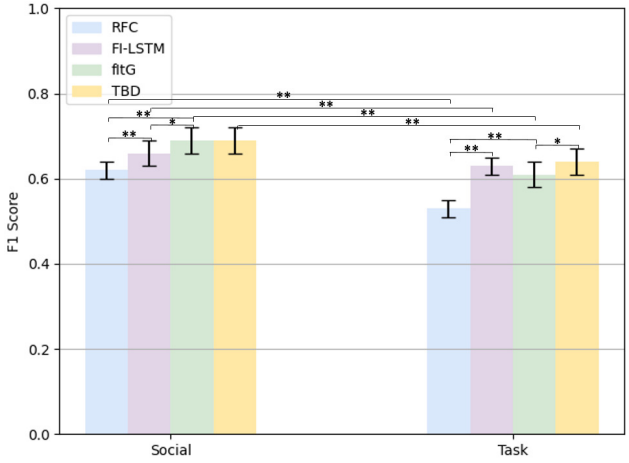


Figure 3: F1-scores of the Social and Task dimensions of cohesion for the RFC_MD (in blue), FI-LSTM_MD (in purple), fItG_MD (in green) and the TBD (in yellow) models. P-values of significant differences are displayed for each dimension and between dimensions, for each model.

These results highlight the benefits for a model to integrate temporality and to learn higher representations of both individuals and group to predict the dynamics of cohesion, especially for the Social dimension. Indeed, for this dimension, the fItG_MD significantly outperformed the other 2 models, confirming the importance of modeling the group at both individual and syntality levels.

5.2 Multilabel vs a Social Sciences inspired approach

Concerning the dynamics of Task dimensions, TBD obtains a significant improvement in the performances with respect to the fItG_MD (p = .028) (see also Table 4). Indeed, it reaches a F1-score of 0.64 ±0.03, improving the fItG_MD performances by 0.03. Such an improvement means that TBD learnt new behavioral patterns that globally improved predictions. TBD, however, does not significantly improve the fItG_MD in all the tasks (see Figure 4): it considerably enhanced Task 2 from a F1-score of 0.55 ±0.11 to 0.64 ±0.09 (p = .038) whereas Task 5 has a significant decrease in performances passing from a F1-score of 0.78 ±0.03 to 0.74 ±0.04 (p = .004). In Task 2, group members had to concentrate on solving problems on their own, limiting their movements and interactions while in Task 5, they had to collaborate to agree to a solution. A possible explanation for this trade-off in performances is that, for a similar label (e.g., decrease), individual and group behaviors are extremely different in these 2 tasks.

These results confirm that integrating the interplay of Social and Task dimensions is beneficial for the prediction of the dynamics of Task cohesion. Since both models take into account the dynamics of cohesion and model the group at both individual and syntality levels, they differ by the way they integrate the Social and Task interplay. Motivated by the Social Sciences theories claiming that, in groups of friends, Social cohesion is more salient than Task cohesion [47] and that the Social dimension first emerges [19] and creates a favorable environment to consolidate Task cohesion [43], TBD takes advantage of a transfer learning approach to integrate this dimensions interplay. Such an approach enables TBD to use the Social representation of cohesion previously learnt by the pre-trained model to optimize the prediction of the Task dynamics. Furthermore, by enabling the retraining of the pre-trained model during the TBD training phase, we also integrate the impact of Task cohesion on Social cohesion.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented an approach to exploit insights from Social Sciences to build computational models of cohesion taking into account the interplay between its Social and Task dimensions over time. In particular, we described TBD, a DNN architecture that, in addition to that, also incorporates temporal dependencies by integrating, amongst others, LSTM layers and models individual as well as collective contributions. TBD was evaluated vs 3 baselines modeling cohesion in different ways. Our architecture allows us to reach a significantly better F1-score for Task cohesion, that as stated in literature is particularly hard to predict (e.g., [37]).

Cohesion is a complex social phenomenon and this work provides an approach to use some insights from Social Sciences to improve computational models. It is, indeed, not exhaustive and has some limitations. Firstly, all of the models are predicting the dynamics of cohesion for the whole interaction once all the thin slices are processed. In the future, we aim to design a model that would predict the dynamics of cohesion at each task, relying solely on the thin slices of the previous and/or current task(s) instead of the whole interaction. Such a model would help leaning toward the development of a "real-time" application. Furthermore, TBD takes advantage of Social cohesion to predict Task cohesion. It would also be interesting to develop a different TBD architecture to explore whether and how Task cohesion improves Social cohesion prediction. Although evidence towards this interplay is, to the best of our knowledge, under-investigated in Social Sciences literature, it could help to have a better understanding of cohesion. In this work, DNNs were designed to integrate a pre-fixed number of person. Here, we tested the architectures on groups of 3. Adding a new person to a group would imply retraining the models. In the future, an architecture able to dynamically self-adapt to various sizes of groups should be developed. In this paper, we focused on the Social and Task dimensions of cohesion only. A new and open challenge will be to build computational models that can also take into account other dimensions (e.g., group pride) and their interplay.

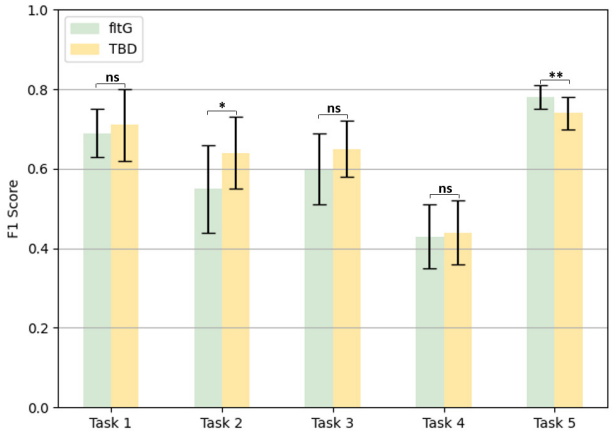


Figure 4: F1-scores per task and for the Task dimension of the fItG_MD (in green) and the TBD (in yellow). The significance level is indicated for each task (ns stands for not significant).

Table 4: F1-scores of the Social and Task dimensions for the fItG_MD and the TBD models. As TBD is using the fItG for predicting the dynamics of the Social dimension, performances are similar for this dimension.

	F1 scores ±std		
	Social	Task	
	fItG_MD & TBD	fItG_MD	TBD
T1	0.52 ±0.08	0.69 ±0.06	0.71 ±0.09
T2	0.59 ±0.12	0.55 ±0.11	0.64 ±0.09
T3	0.61 ±0.06	0.60 ±0.09	0.65 ±0.07
T4	0.88 ±0.03	0.43 ±0.08	0.44 ±0.08
T5	0.84 ±0.05	0.78 ±0.03	0.74 ±0.04
Average	**0.69 ±0.03**	**0.61±0.03**	**0.64 ±0.03**

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] Nalini Ambady and Robert Rosenthal. 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of personality and social psychology* 64, 3 (1993), 431–441.

[2] Aristotle. 350BCE. *Metaphysics*. by W. D. Ross, Translated.

[3] Kurt W. Back. 1951. Influence through social communication. *The Journal of Abnormal and Social Psychology* 46, 1 (1951), 9–23.

[4] Daniel J Beal, Robin R Cohen, Michael J Burke, and Christy L McLendon. 2003. Cohesion and Performance in Groups: A Meta-Analytic Clarification of Construct Relations. *Journal of Applied Psychology* 88, 6 (2003), 989–1004.

[5] Kenneth A Bollen and Rick H Hoyle. 1990. Perceived Cohesion: A Conceptual and Empirical Examination. *Social Forces* 69, 2 (1990), 479–504.

[6] Albert V Carron and Lawrence R Brawley. 2000. Cohesion: Conceptual and Measurement Issues. *Small Group Research* 31 (2000), 89–106.

[7] 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.

[8] Raymond B Cattell. 1948. Concepts and methods in the measurement of group syntality. *Psychological Review* 55, 1 (1948), 48–63.

[9] Eleonora Ceccaldi, Nale Lehmann-Willenbrock, Erica Volta, Mohamed Chetouani, Gualtiero Volpe, and Giovanna Varni. 2019. How unitizing affects annotation of cohesion. In *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.

[10] Edward Clarkson, Jason A. Day, and James D. Foley. 2006. *An Educational Digital Library for Human-Centered Computing*. Association for Computing Machinery, 646–651.

[11] 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]

[12] Kenneth L Dion. 2000. Group Cohesion: From Field of Forces to Multidimensional Construct. *Group Dynamics: Theory, Research, and Practice* 4, 1 (2000), 7–26.

[13] Metehan Doyran et al. 2021. MUMBAI: multi-person, multimodal board game affect and interaction analysis dataset. *Journal on Multimodal User Interfaces* (2021), 1–19.

[14] Steven J. Kamper et al. 2010. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of Clinical Epidemiology* 63, 7 (2010), 760–766.

[15] Florian Eyben 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.

[16] Daniel Gatica-Perez, Iain McCowan, Dong Zhang, and Samy Bengio. 2005. Detecting group interest-level in meetings. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1. IEEE, 489–492.

[17] James Griffith. 1988. Measurement of Group Cohesion in U.S. Army Units. *Basic and Applied Social Psychology* 9, 2 (1988), 149–171.

[18] Rebecca Grossman, Sarit B Friedman, and Suman Kalra. 2017. Teamwork processes and emergent states. *The wiley blackwell handbook of the psychology of team working and collaborative processes* 42 (2017), 243–269.

[19] Rebecca Grossman, Zachary Rosch, David Mazer, and Eduardo Salas. 2015. What Matters for Team Cohesion Measurement? A Synthesis. *Research on Managing Groups and Teams* 17 (2015), 147–180.

[20] Nils Y. Hammerla, Shane Halloran, and Thomas Ploetz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables. arXiv:1604.08880 [cs.LG]

[21] Anjali Hans and Emmanuel Hans. 2015. Kinesics, haptics and proxemics: Aspects of non-verbal communication. *IOSR Journal of Humanities and Social Science* 20, 2 (2015), 47–52.

[22] 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.

[23] Hanbyul Joo et al. 2019. Panoptic Studio: A Massively Multiview System for Social Interaction Capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, 1 (2019), 190–204.

[24] Gary King and Langche Zeng. 2001. Logistic regression in rare events data. *Political analysis* 9, 2 (2001), 137–163.

[25] Andrey Kolmogorov. 1933. Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.* 4 (1933), 83–91.

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

[27] Kurt Lewin. 1946. Behavior and development as a function of the total situation. In *Manual of child psychology*. John Wiley & Sons Inc, 791–844.

[28] Kurt Lewin. 1951. *Field theory in social science: selected theoretical papers (Edited by Dorwin Cartwright)*. Harpers.

[29] Robert G. Lord, John F. Binning, Michael C. Rush, and Jay C. Thomas. 1978. The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior. *Organizational Behavior and Human Performance* 21, 1 (1978), 27–39.

[30] Albert J Lott and Bernice E Lott. 1965. Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychological Bulletin* 64, 4 (1965), 259–309.

[31] Nina Cristina Magpili and Pilar Pazos. 2018. Self-managing team performance: A systematic review of multilevel input factors. *Small Group Research* 49, 1 (2018), 3–33.

[32] 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.

[33] 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.

[34] Michelle A Marks, John E Mathieu, and Stephen J Zaccaro. 2001. A Temporally Based Framework and Taxonomy of Team Processes. *The Academy of Management Review* 26, 3 (2001), 356–376.

[35] Iain McCowan et al. 2005. The AMI meeting corpus. In *Proceedings of the 5th international conference on methods and techniques in behavioral research*, Vol. 88. Citeseer, 100–104.

[36] Philipp Müller, Michael Xuelin Huang, and Andreas Bulling. 2018. Detecting Low Rapport During Natural Interactions in Small Groups from Non-Verbal Behaviour. In *23rd International Conference on Intelligent User Interfaces*. Association for Computing Machinery, 153–164.

[37] Marjolein C Nanninga, Yanxia Zhang, Nale Lehmann-Willenbrock, Zoltán Szlávik, and Hayley Hung. 2017. Estimating Verbal Expressions of Task and Social Cohesion in Meetings by Quantifying Paralinguistic Mimicry. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. Association for Computing Machinery, 206–215.

[38] Roland Nilsson, José M Pena, Johan Björkegren, and Jesper Tegnér. 2007. Consistent feature selection for pattern recognition in polynomial time. *The Journal of Machine Learning Research* 8 (2007), 589–612.

[39] Xinglong Qu, Shun Kurokawa, and The Anh Han. 2021. Social cohesion V.S. task cohesion: An evolutionary game theory study. arXiv:2101.06961 [physics.soc-ph]

[40] Fabien Ringeval, Erik Marchi, Charline Grossard, Jean Xavier, Mohamed Chetouani, David Cohen, and Björn Schuller. 2016. Automatic analysis of typical and atypical encoding of spontaneous emotion in the voice of children. In *Proceedings of the 17th Annual Conference of the International Speech Communication Association (ISCA)*. ISCA, 1210–1214.

[41] Lisa Rosh, Lynn R Offermann, and Rhonda Van Diest. 2012. Too close for comfort? Distinguishing between team intimacy and team cohesion. *Human Resource Management Review* 22, 2 (2012), 116–127.

[42] Eduardo Salas, Rebecca Grossman, Ashley M Hughes, and Chris W Coultas. 2015. Measuring Team Cohesion: Observations from the Science. *Human Factors* 57, 3 (2015), 365–374.

[43] Jamie B Severt and Armando X Estrada. 2015. On the Function and Structure of Group Cohesion. In *Team Cohesion: Advances in Psychological Theory, Methods and Practice*. Vol. 17. Emerald Group Publishing Limited, 3–24.

[44] Nickolay Smirnov. 1948. Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics* 19, 2 (1948), 279–281.

[45] Daniel Szafir and Bilge Mutlu. 2012. Pay Attention! Designing Adaptive Agents That Monitor and Improve User Engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 11–20.

[46] Lisa Torrey and Jude Shavlik. 2010. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI global, 242–264.

[47] Bruce W Tuckman and Mary Ann C Jensen. 1977. Stages of small-group development revisited. *Group & Organization Studies* 2, 4 (1977), 419–427.

[48] Michael Wainberg, Babak Alipanahi, and Brendan J Frey. 2016. Are random forests truly the best classifiers? *The Journal of Machine Learning Research* 17, 1 (2016), 3837–3841.

[49] Fabian Walocha, Lucien Maman, Mohamed Chetouani, and Giovanna Varni. 2020. Modeling Dynamics of Task and Social Cohesion from the Group Perspective Using Nonverbal Motion Capture-Based Features. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*. Association for Computing Machinery, 182–190.

[50] 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.

[51] Yanxia Zhang et al. 2018. TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)* 2, 3 (2018), 1–22.