# Modeling the Interplay Between Cohesion Dimensions: a Challenge for Group Affective Emergent States

Lucien Maman, Nale Lehmann-Willenbrock, Mohamed Chetouani, Laurence Likforman-Sulem *Senior Member, IEEE,* and Giovanna Varni

**Abstract**—Emergent states are temporal group phenomena that arise from collective affective, behavioral, and cognitive processes shared among the group's members during their interactions. Cohesion is one such state, mainly conceptualized by scholars as affective in nature, and frequently distinguished into the two dimensions social and task cohesion. Whereas social cohesion is related to the need of belonging to a group, task cohesion is related to the group's goals and tasks. In this paper, we emphasize the importance of behavioral interaction dynamics for predicting cohesion's dynamics. Drawing from Social Science insights, we investigate the interplay between social and task cohesion to predict their dynamics across group tasks from nonverbal behavioral features. Three computational architectures exploiting transfer learning are presented. Transfer learning capitalizes on information learnt by a model for a specific dimension to predict the dynamics of the other dimension. Results show that integrating the influence of social cohesion for predicting dynamics of task cohesion outperforms state-of-the-art. For predicting dynamics of social cohesion, a model integrating the reciprocal impact of social and task cohesion significantly improves performance with respect to the state-of-the-art, as well as compared to a model only integrating the impact of task cohesion on dynamics of social cohesion.

**Index Terms**—Group Emergent States, Cohesion, Group Dynamics, Multimodal Interaction, Transfer Learning

## 1 Introduction

Implicit and explicit interactions among the members of a group for coordinating their actions and intents to achieve objectives shape affective, cognitive and behavioral group processes and outcomes [1]. As a result of such interactions, the literature categorizes some group phenomena as *emergent states* (e.g., [1], [2], [3], [4]) that come into existence due to group members' behaviors expressed over the course of dynamic group interactions. It follows from this conceptualization that, in order to understand how emergent group states come about, we need to study the underlying behavioral group interactions. Traditionally emergent states are captured through self-report surveys, which are relatively easy to obtain but only allow very limited insights into the dynamics of such states [5], [6]. Obtaining a fine grained temporal resolution of emergent states would indeed requires the administration of the surveys at many times during interaction, as well as the need of disguising these measurement tools to make them unobtrusive. Recent efforts by scholars in Computer Science to develop machines able to engage humans in more effective activities over the course of dynamic group interactions point to computational approaches as alternative measures of emergent states. Computational approaches can overcome limitations of traditional survey measures, by allowing insights into the actual behavioral dynamics of emergent group processes.

Group cohesion is a multidimensional emergent state and it is the one mainly adopted as a test bed for investigating such computational approaches (see Section 2.2). Cohesion is broadly defined as "the tendency for a group to stick together and remain united to pursue goals and/or affective needs" [7]. As this definition suggests, in addition to the seminal article by Marks and colleagues [2] as well as other studies (e.g. [8], [9], [10]), cohesion is an affectively laden construct, or an affective emergent group state. The specific affective nuance of cohesion, combined with the development of behavioral computational approaches, corroborates the need for the Affective Computing community to include cohesion and other affect-laden group processes among its research topics. Importantly however, while cohesion is an affectively laden construct, social interaction behaviors matter greatly to the emergence of group cohesion (for a detailed discussion, see [11]).

In this paper, we present a behavioral computational approach to the dynamics of cohesion, focusing on the facets of its instrumental function, i.e., *social* cohesion and *task* cohesion, and their interplay. We adopted the GAME-ON dataset conceived to investigate the dynamics of cohesion in group of friends [12]. In such a setting, as reported in the literature, we can reasonably expect that cohesion has already emerged and less volatile than in newly formed groups. Thus, we focus on the dynamics of a "well-established" cohesion and


- *L. Maman, L. Likforman-Sulem are with LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France. E-mails: {lucien.maman, laurence.likforman}@telecom-paris.fr*
- *G. Varni is with Department of Information Engineering and Computer Science (DISI), University of Trento, Italy; she carried out the work in the paper while she was with LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France. E-mail: giovanna.varni@unitn.it*
- *N. Lehmann-Willenbrock is with Department of Industrial/Organizational Psychology, University of Hamburg, Germany. E-mail: nale.lehmann-willenbrock@uni-hamburg.de*
- *M. Chetouani is with Institute for Intelligent Systems and Robotics, Sorbonne University, CNRS UMR7222, Paris, France. E-mail: mohamed.chetouani@sorbonne-universite.fr*


*Manuscript received January XX, 2022.*

not on the dynamics of the emergence of cohesion. Concretely, we conceived three Deep Neural Networks (DNN)-based architectures for predicting the interlaced dynamics of social and task cohesion, that is, they exploit knowledge learnt on one dimension of cohesion for predicting the dynamics of the respective other one. Dynamics here refers to changes in cohesion between a pair of consecutive tasks, and its prediction is formulated as a binary classification problem, that is, decrease vs not-decrease (not-decrease including both stability and increase). The first two architectures, called *Transfer Between Dimensions* (TBD), largely extend the work presented in [13]. Exploiting a transfer learning approach, they take advantage of the information learned by a model for a specific dimension to predict the dynamics of the other one. More specifically, they use a pre-trained model for predicting dynamics of social cohesion to predict task cohesion dynamics and vice-versa. The third architecture, called TBD-*Reciprocal Impact* (TBD-RI), is built on top of the first two, and not only does it take advantage of the information learned by each pre-trained model, but also integrates the reciprocal impact between social and task cohesion. Moreover, following suggestions from Social Sciences (e.g., [6]), all these architectures: i) integrate time dependencies (i.e. they look at the history of what the group performed), and ii) take into account the contribution of each group member as well as the performance of the group as a whole. The performances of the models are evaluated against a state-of-the-art model predicting the dynamics of social and task cohesion in a multilabel setting [14].

The remainder of this paper is as follows: Section 2 conceptualizes cohesion as a behavioral team construct, drawing from recent Social Science research, and also reviews existing computational studies. Section 3 explains the motivations behind the design of the DNN-based architectures. In Section 4, the experimental setting is presented by illustrating the dataset, the labeling strategy, and the multimodal nonverbal features used. Next, Section 5 describes the architectures, whereas their evaluation is presented in Section 6. Then, Section 7 reports the results, and Section 8 discusses them.

## 2 Background and Related Work

### 2.1 Cohesion: a multidimensional emergent state

Early definitions of cohesion were based on the force field theory, considering individual and group actions as influenced by external forces [15]. Subsequently, various definitions (e.g., [5], [16], [17], [18], [19]) as well as theoretical models and frameworks of cohesion emphasize its affective component (e.g., [7], [20], [21], [22]), rather than its behavioral underpinnings in group and team interactions. For example, Kozlowski and Chao described cohesion as an "affectively loaded" emergent state [5]. Similarly, Maynard *et al*. stated that cohesion is "imbued by affective and emotional forces" [19]. Some authors explicitly associate social bonds and the sense of belonging to a group (e.g., [17], [23]). For example, Beal *et al.*, [17] identified two components, or dimensions, of cohesion, that is, *Interpersonal Attraction* and *Group Pride*. Interpersonal Attraction refers to the shared liking among the group members, while Group Pride is conceptualized as the shared sense of honor derived from being a member of a group. Of note, while these definitions emphasize the affective connotations of cohesion, we still maintain that behavioral group interactions will determine the various levels of cohesion experienced by group members. These include task-focused interaction behavior as well as relational interaction [24]. Whereas communication scholars typically view cohesion as a relational construct (e.g., [25]), we also foresee that relational and task-based component of cohesion are intertwined. Affect can also manifest when group members engage in high-quality social working relationships, hence creating a positive working environment to accomplish group goals and tasks. In this way, affect is more related to the goal- and task-based activities of the group, as expanded in one of the most popular cohesion models developed by Carron and Brawley [7]. In this model, cohesion is defined by the *Individual Attraction to Group* and the *Group Integration* components that are both divided into the *social* and *task* dimensions. For both components, the authors state that all the reasons that would motivate a group member to remain in a group and to stay united could manifest through the social and task dimensions of cohesion. This notion also hints at the possibility that the two dimensions of cohesion are entangled, and may exert reciprocal influences on each other. Building on and integrating previous definitions and models, Severt and Estrada introduced a multidimensional theoretical framework to categorize the structural and functional properties of cohesion [22]. According to this framework, cohesion serves two main functions that each comprise two separate but interrelated dimensions. The first function refers to the emotional benefits group members can experience in a group. It is composed of the *Interpersonal* and *Group Pride* dimensions. The Interpersonal dimension refers to the friendships bonds that develop over time. The emergence of Group Pride is driven by the tendency of individuals to identify with a successful group as well as the desire to define one's role within it. The second function of cohesion involves all the aspects of cohesion that highlight the goal- and task-based activities of the group. It is organized in the social and task dimensions. Furthermore, as in [20], the authors distinguish two levels at which cohesion can be observed: *horizontal* and *vertical*. Horizontal cohesion concerns relations among peers, whereas vertical cohesion implies hierarchy referring to the relationships between a member of authority and a subordinate within the group context.

In this study, following the theoretical framework of Severt and Estrada, we computationally investigate the temporal dynamics of social and task dimensions of cohesion, with a particular focus on their interplay. Specifically, our study helps pinpoint the temporal fluctuation of social and task cohesion levels and their reciprocal influence over time. Furthermore, we investigate cohesion at the horizontal level, since it aligns with the modern shift towards holacracy organizational structures and self-managed teams [26].

### 2.2 Cohesion and computing

Despite the variety of definitions, theoretical models and frameworks of cohesion offered by the literature, computational studies on cohesion are still scarce and neglect how its dimensions are interrelated between them. More specifically, first studies investigated which behavioral features

concur to model and predict cohesion as an emergent state. More recently, the researchers acknowledged the relevance of cohesion as an affective emergent state and explicitly integrate affect in their computational models.

Among the first class of studies, Hung and Gatica-Perez, in their pioneering work, extracted a set of nonverbal audio and visual features to capture the dimensions of cohesion related to its observable aspects (e.g., rapport, involvement or mimicry) and affect [27]. They explored what are the most important features for inferring cohesion as a whole (i.e., without distinguishing between its dimensions) using supervised classifiers (e.g., Support Vector Machine). Their findings show, for example, the relevance of turn-taking-related features. Nanninga *et al.*, extended this work, integrating pairwise and group features related to the alignment of para-linguistic speech behavior [28]. Para-linguistic mimicry cues are, indeed, among the most important ones in signaling emotions [29]. In their study, they modeled, through a Gaussian Naive Bayes classifier, the social and task dimensions separately, and they predicted social dimension in a more accurate way than task dimension. More recently, Walocha *et al.* also investigated what are the most relevant movement features for predicting the dynamics of the social and task dimensions. They trained and evaluated a Random Forest model fed with features extracted from motion capture data and using labels of cohesion dynamics built from self-assessments. The relevance of the features was assessed via their Shapley values. The results show that proxemics and kinesics features are the most successful at predicting social and task dynamics of cohesion. The findings of all these studies, however, are often constrained by the type of machine learning model used. Therefore, they might not generalize to all the variety of models.

Among the computational approaches that explicitly integrate affect in models' architecture, Dhall and colleagues set up with a bench-marking platform to jointly investigate how to address cohesion and emotion within the context of the EmotiW challenge [30]. As part of this challenge, researchers implemented various DNNs to predict group cohesion from images and videos with (e.g., [31], [32], [33]) and without (e.g., [34]) explicitly collecting information about emotions. These works, however, only considered cohesion as a whole, without distinguishing between its dimensions. Lately, Maman et al., presented two DNN architectures that implement a multitask learning approach to jointly predict the social and task cohesion dynamics as well as the valence of group emotion [14]. There, affect is taken into consideration in both the features and the architecture. The DNN architectures are inspired by the Top-down and Bottom-up group emotion approaches described in [35]. The result show that a Bottom-up approach significantly improves the prediction of the dynamics of task cohesion. All of these studies, however, integrate the relationship between cohesion and emotions in their architectures as a proxy for modeling the affective role played by cohesion, neglecting its innate affective function.

The architectures presented in this paper leverage both these two classes of studies, and explore the interplay between the social and task dimensions of cohesion grounding on Social Sciences' insights.

## 3 The interplay of social and task cohesion

Although scholars in Social Sciences clearly state that the cohesion's dimensions interplay somehow and somewhere over time, it is less clear if and how this might occur empirically. Some authors argue that social cohesion emerges first and may impact the development of task cohesion (e.g., [36], [37]). Other ones affirm that, especially at an early stage of group formation, task cohesion might emerge before the social one, and it could be seen as a shared experience auspicious to group bonding (e.g., [38]). These two opposite points of view might hold depending on many factors (e.g., the nature of the group members and group's goals). In their work, Severt and Estrada, indeed, highlight that not every group exploits each dimension of cohesion [22]. Moreover, Grossman et al. state that once social cohesion appeared followed by task cohesion, after a while, a dynamical reciprocal adjustment between the two dimensions occurs, at the expense of social cohesion [37]. To the best of our knowledge, the two last aforesaid studies are the only ones mentioning such an adjustment, opening a third way to study the interplay between the social and task dimensions of cohesion. These different perspectives suggest that we need to investigate both possible directions of influence between the task and social cohesion dimensions.

### 3.1 From social cohesion to task cohesion

Early work by Tuckman on small group development suggests that cohesion is part of the life cycle of a group and that the social dimension of cohesion develops first [36]. Empirical work confirmed and extended Tuckman's hypothesis (e.g., [39], [40]) stating that groups go through the stages of *forming*, *storming*, *norming*, *performing*, and, finally, *adjourning* [41]. During the forming, group members develop social bonds and get to know each other, while, in the storming, they start learning about each others' strengths and weaknesses, leading to the definition of their roles. Such a categorization of the different stages of a group encourages to consider the social dimension as a potential driver for the task one. Moreover, Carron and Brawley state that all dimensions are not equally present across groups and that some dimensions might be more salient depending on the developmental phase of the group (e.g., newly formed groups), and the specific interaction setting (e.g., a meeting) [7]. In addition, the influence of a dimension on another is likely to change gradually over time. In their paper, they also conclude that, in particular contexts (e.g., in social groups), social cohesion would be more salient. In [37], Grossman and colleagues support the predominance of social cohesion in social groups and argue that social cohesion emerges first, and sets the stage for task cohesion, which develops later on. Lending further support to the notion that social cohesion breeds task cohesion, Severt and Estrada [22] advanced that social cohesion facilitates flexible and constructive relationships in groups and teams, hence, promoting task cohesion. While this relationship does not imply causation, previous studies (e.g., [22], [37]) converge on the impact of social cohesion on task cohesion across different group development stages and settings (e.g., number of persons, context).

### 3.2 From task cohesion to social cohesion

While the path from social cohesion to task cohesion may be more intuitive from a group developmental point of view, the other path (i.e., task cohesion influencing social cohesion) may also occur. Prior theorizing has hinted at the possibility that task cohesion might emerge earlier in a group's developmental trajectory, before group bonding and relationship formation come into play and create shared experiences of social cohesion [38]. In earlier stages of team development, task aspects can be more salient than the social ones, which may require an extended period of interaction [7]. Empirical work indicates support for the notion of task aspects potentially influencing social cohesion. A study of youth athletes showed that members of task-focused teams report personal enjoyment and friendship development [42]. Similarly, a study of teams of male college athletes showed that a task-involving team climate predicts aspects of social cohesion [43]. The authors discuss that a task-involving climate can help reduce social barriers, foster interdependence, and trigger positive social interactions, which paves the way for social cohesion. While it remains to be seen whether these findings extend to other types of groups with a more heterogeneous gender distribution, we interpret these earlier results in terms of a possible link from task to social cohesion. Such a relationship also does not imply causation. However, it may be more subject to fluctuations over time [7], highlighting the importance of observing the impact from task cohesion to social cohesion at various stages of group development.

## 4 Experimental Setting

### 4.1 Dataset

We adopted the GAME-ON dataset [12], specifically designed to study the dynamics of social and task cohesion over time. It consists of more than 11 hours of multimodal data (i.e., video, motion capture, and audio recordings) from 15 groups of friends playing an escape game scenario. Each group is composed of three different members, for a total of 45 persons (69% of participants identified themselves as female and 31% identified as male). Five triads are composed of female members only, ten are composed of female and male members. No triads is composed of male members only. The participants' ages ranged from 21y to 33y (M = 25.3y, SD = 3.1y).

The escape game comprised five tasks designed ad hoc to elicit variations of cohesion along the two dimensions (i.e., increase or decrease with respect to the previous task). In *Task 1* (T1), group members competed to find a key and a box hidden in the room. In *Task 2* (T2), the group had to resolve a set of enigmas. Once dispatched to everyone, each group member had to solve enigma on their own. In *Task 3* (T3), each member had to solve a complex problem in a limited time that required information from the other members. In *Task 4* (T4), the group had to guess the signification of an unusual object, while in *Task 5*(T5), they had to present how all the hints were interrelated to escape the room. The average duration of the game was 35min 30s (SD = 4min 10s).

TABLE 1
Expected variations of social and task cohesion across tasks. In the first column there are the transitions between two consecutive tasks ($T_i$ is one among the five tasks, with $T_0$ the beginning of the game); in the second column the tasks' duration (average and standard deviation); in the third column the expected variations of social and task cohesion: '-' and '+' stand for decrease and not decrease, respectively. The '-' and '+' in bold and double quoted highlight variations that were not confirmed by the groups' self-reports.

| | | Expected variations in cohesion | |
|---|---|---|---|
| **Transition** | **Avg Duration ±std** | **Social** | **Task** |
| **T0 - T1** | 8min 29s ±1min 34s | - | - |
| **T1 - T2** | 7min 33s ±1min 09s | - | **+** |
| **T2 - T3** | 6min 27s ±1min 28s | + | **-** |
| **T3 - T4** | 5min 48s ±1min 55s | + | + |
| **T4 - T5** | 7min 13s ±1min 52s | + | + |

The dataset also includes, for each group, repeated self-assessments of cohesion provided by every member. Cohesion was assessed at the beginning and at the end of each task through a slightly modified version of the Group Environment Questionnaire (GEQ) [44], a well-established questionnaire composed of 18 items with a 9-point Likert scale answering format. This questionnaire has already been used in various studies to measure the social and task dimensions of cohesion (e.g., [45], [46]). The GEQ version adopted in GAME-ON consisted of eight items related to the task dimension and six items related to the social dimension, respectively. Some items were adapted to the context of the escape game without changing the valence nor the grammatical construct, whereas two items were replaced since they are close enough to the originals and more suited to the context (see [12] for details). For each member and for each task, a cohesion score was computed for the social and task dimensions of cohesion by summing the items associated to the corresponding dimension. The authors showed that, except for the T1-T2 and T2-T3 transitions of task cohesion, the expected variations of cohesion for both dimensions are confirmed (see Table 1).

### 4.2 Multimodal nonverbal features

Previous works show that nonverbal communication is a more powerful predictor of group cohesion than verbal behavior (e.g., [47], [48]). For that reason, we extracted 84 nonverbal multimodal features characterizing social interaction from the GAME-ON's motion capture data and the audio recordings. The features were extracted for individuals (I) as well as for the whole group (G) over the last two minutes of each task and in consecutive time windows lasting 20s. The choice to focus on the last two minutes of each task was motivated by the use, in this work, of the self-assessments on cohesion provided by the group members. As reported in several studies, indeed, self-assessments collected through questionnaires are likely influenced by the last recalled behavior (e.g., [49], [50]). The duration of the time windows (i.e., 20s) is grounded on previous work on group interaction (e.g., [51]) and cohesion perception (e.g., [52]). In the remainder of this section, statistics are computed on these time windows unless differently specified. GeMAPS features are extracted on sub-windows having sizes as recommended in [53].

#### 4.2.1 Motion capture features

*Proxemics*

Proxemics is the study of how humans use and structure space around them [54]. Previous studies show its relevant role in nonverbal communication and social interaction (e.g., [55]). As empirically demonstrated by Ashton et al., we expect groups that are standing closer together to not interpret the presence of others as invading, meaning they have a stronger social bond to each other and to trigger positive affective reactions [56]. The following proxemics-related features were extracted:

- *Histogram of the interpersonal distance (G)*. Interpersonal distances are computed frame by frame as the Euclidean distance between the projection of the spines of two group members over the transverse plane. This procedure is repeated for each pair. Then, according to [54], they are clustered as follows: public space ($>$ 3.6 m), social space (in 3.6 m and 1.2 m) or personal space ($<$ 1.2 m), respectively.
- *Maximum of the interpersonal distances (G)*[1]. Based on the interpersonal distances computed previously, the maximum distance among the three pairs of hips at each frame is selected.
- *Distances from the group barycenter (I)*[1]. The spine of each group member is projected on the transverse plane. The group barycenter on such plane is computed as the barycenter of the triangle having as vertexes the three spine's projections. Finally the Euclidean distance of such projections and the barycenter is computed.
- *Total distance traveled (I)*, computed as the length of the trajectory covered by the spine's projection of each group member on the transverse plane.
- *Time in F-formation (G)*, it is the amount of time during which a group make a circular or a semi-circular F-formation [57]. We focused on the circular and semi-circular ones because they are indicative of a shared-interest in the interaction [57]. To automatically detect these F-formations, a cone is computed from the chest of each member to approximate the area where the group members' attention is directed. An F-Formation is detected when the cones of every group member intersect.

*Kinesics*

Kinesics concerns the study of how humans communicate using posture, gesture, stance, and movement [58]. Kinesics features may indicate active engagement in the task and thus are expected to have a positive impact on predicting cohesion [59]. We extracted the following features:

- *Kinetic energy (I)*[1], computed as the sum of the translational and rotational kinetic energies of the whole body of each member. For the sake of simplicity, masses and moments of inertia were taken equal to one.
- *Synchrony among kinetic energies (G)*, computed as the S-Estimator of the kinetic energies of each member. S-Estimator is a measure of synchronization exploiting the normalized eigenvalues of the correlation matrix of multiple signals [60].
- *Group amount of motion (G)*[1], computed as the standard deviation over 1s of trajectory followed by the projection of each group member's chest over the transverse plane. The average among these three values is then computed to get a group feature at each second, resulting in 20 values for the time window.
- *Group's amount of hand movements while not moving (G)*[1], computed as the standard deviation of the 3D displacement of each member's left and right hands. A member is considered as not moving if its hip position over the transverse plane did not exceed 50cm over one second. At each second, the mean between both hands movement of the three group members is computed, resulting in 20 values for the time window. Hands movements are a vector for specific emotions communication [61] and might also be indicative of the group engagement in the task.
- *Posture expansion (I)*: we computed the variations of the volume of the bounding box (i.e., the smallest box containing the body joints) [62]. Moreover, we computed the variations of the area of the bounding rectangles on the frontal and transverse planes. Posture expansion is expected to be related to dominance and hierarchy, small differences and big overall expansion being positively correlated to social cohesion [63] and emotions [64] and relevant for studying other affective phenomena (e.g., stress detection [65])
- *Touch's duration (G)*, computed as the overall amount of time a hand of a group member is touching the upper body of another one. Here, a touch is detected when the sensor located on the hand of a member is less than 15cm away from a sensor on the upper body of another one. Signaling by touch can communicate task-related information as well as convey social status and emotions [66].

Kinetic energy and posture expansion were filtered using a Savitzky-Golay filter [67] with a polynomial order of five and a coefficient of three to reduce noise.

#### 4.2.2 Audio features

*GeMAPS features*

Features of the Geneva Minimalistic Acoustic Parameter Set (GeMAPS) [53] were extracted using the OpenSmile software [68]. We chose GeMAPS since it has been successfully used in many affect-related prediction tasks (e.g., [69], [70]). Moreover, it has been proven relevant for predicting various other social phenomena (e.g., amusement, interest) [71]. GeMAPS includes the following features, for which the mean was applied over each time window.

- *Frequency related (I)*: Pitch, Jitter, F1, F2 and F3 frequencies and F1 bandwidth. Such features are particularly relevant for describing vocal affective expressions and, in particular, anger and sadness [71].
- *Energy and amplitude related (I)*: Shimmer, Loudness and Harmonic-to-Noise Ratio. These features are pertinent to detect, for example, stress [72].
- *Spectral (balance) (I)*: Alpha Ratio, Hammarberg Index, Spectral Slope (0-500 Hz and 500-1500 Hz), F1,

1. Mean, std, min, max and skewness statistics were applied over all the values computed within each 20s time window

F2 and F3 relative energies, and Harmonic Difference (H1-H2 and H1-A3). They had been successfully used for the detection of angry speech [73], and they are also important for vocal valence and arousal [71].

*Conversational features*

Previous work on automated detection of cohesion shows the relevance of taking into account conversational features (e.g., [27]). First, a speech matrix was computed using the voice activity detector (VAD) from Opensmile [74]. Then, the following features were extracted from this matrix:

- *Average turn duration (G)*, it is the average duration of all the turns occurring during a group interaction. A turn is considered over when a member stops speaking for at least one second. In an extremely involved conversation, turns duration of each participant is theorized to be approximately equal [27]. Also, we would expect that in highly cohesive groups, turns will tend to be shorter as everyone would freely contribute to the conversation.
- *Time of overlapping speech (G)*, it is the total time for which at least two members speak simultaneously. Overlapping can be symptomatic of conflict or be a sign of engagement between people [75].
- *Total speaking time (I)*, computed for each member as the total time they are speaking. To avoid counting the small utterances, we assume that a member is speaking if she speaks for at least one second.
- *Laughter duration (I)*, automatically extracted using the laugh detector in [76]. Once the laugh is extracted, the total time of laughing is computed for each member. Laughter is, indeed, a highly social phenomenon [77] that is a good indicator of group cohesiveness [78].

### 4.3 Labels

We considered the task of predicting cohesion dynamics as a binary classification problem. Starting from the self-assessments on cohesion rated by each group member, we built labels for decrease vs not-decrease (this one including both stability and increase) as follows. Let's consider the ratings provided by each group member in two consecutive tasks: this results, for each dimension, in six values (two ratings for each of the three members). These ratings were then ranked to limit the potential bias introduced by the inter-member variance. Next, we computed the difference between the ranks of the two consecutive ratings provided by each group member, and we took the average of these differences as the group's cohesion score as shown in Equation 1:

$$GS_{Tx} = \frac{1}{n}\sum_{i=1}^{n} rank_{Tx}^{(i)} - rank_{Tx-1}^{(i)} \quad (1)$$

with $GS_{Tx}$, the group's cohesion score for transition $T_{x-1} - T_x$ with x $\in \{1, 2, 3, 4, 5\}$, $n$ the number of group members (here set to 3), and $rank^{(i)}$ the rank corresponding to the associated rating given by group member $i$. The group's cohesion score indicates whether cohesion decreased or not in a group for a specific dimension. Finally, the score was binarized as follows: a value equal to 0 was assigned when the group score was negative (i.e. a decrease in cohesion occurred), whereas a value equal to 1 was assigned when the group score was equal to 0 or positive (i.e. stability or an increase in cohesion occurred). Overall, this labeling led to an imbalanced distribution for the social dimension (i.e., 75% of "No decrease" labels vs 25% of "Decrease" labels), and to a balanced distribution for the task dimension (i.e., 59% of "No decrease" labels vs 41% of "Decrease" labels). Figure 1 shows the labels distributions of the social and task dimensions, for each task. Such imbalance distributions in the tasks were, however, expected due to how GAME-ON was conceived (see Table 1), and were addressed as described in Section 6.

## 5 DNN-BASED ARCHITECTURES FOR INTEGRATING SOCIAL AND TASK INTERPLAY

### 5.1 Baseline

In this work, we chose the *from Individual to Group* (fItG) architecture [13] (see Figure 2) as a baseline for comparing our architectures. FItG is composed of four components: *Input*, *Individual*, *Group* and *Output*.

Input extracts the multimodal nonverbal features (see Section 4.2). Features computed for each individual are processed by the *Individual* component made of three branches (one for each group member). Each branch consists of a fully connected (FC) layer with a ReLu activation function and 50 units, followed by a Long Short-Term Memory (LSTM) layer with 50 units too. To let the model learn a global representation of an individual, each layer is shared according to Equation 2:

$$\boldsymbol{Y}_i = \phi_i\left(\sum_{j=1}^{n}(W\boldsymbol{X}_j)\right) \quad (2)$$

where $Y_i$ is the output of layer $i$, $\phi_i$, the activation function of the layer $i$, $W$, the matrix of weights common to every group member and $X_j$, the input related to player $j$. As the groups are composed of three individuals, $n$ is set to three. The outputs of these shared layers are then concatenated with the group features and are the input of the Group component.

Group learns the behavior representation at the group level. It is made of a FC layer with a ReLu activation function and 64 units, followed by an LSTM layer with 32 units.

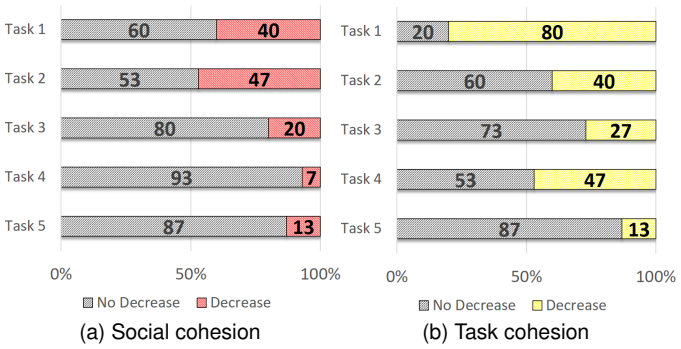



Fig. 1. Labels distributions of social (Figure 1a) and task (Figure 1b) cohesion along the tasks.

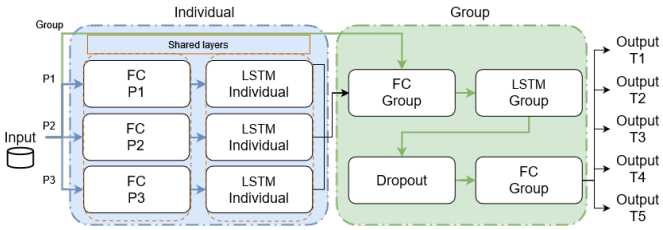



Fig. 2. The fItG model architecture. It has four components: Input, Individual, Group, and Output. Input extracts the features: the individual ones are processed by Individual (in blue), while the group ones are concatenated with the output of Individual into Group (in green). Output predicts the dynamics of social and task cohesion in a multilabel setting.

Next, a Dropout layer with a rate of 0.2 is used, followed by another FC layer with a ReLu activation function and 16 units.

Finally, Output consists of a FC layer with a sigmoid activation function and two units for each task enabling the prediction of the dynamics of social and task cohesion in a multilabel setting.

In total, fItG has 48152 trainable weights. Although this architecture integrates the social and task cohesion interplay, however it does not specify any path from one dimension to the other one. Both dimensions are here tightly related to each other since the overall loss is the sum of the losses from both dimensions and only a final FC layer is differentiating both dimensions.

### 5.2 The TBD architectures

We conceived the *Transfer Between Dimensions* (TBD) architectures taking inspiration by the Social Sciences insights mentioned in Section 3. Due to the contradictory views on which of the two dimensions of cohesion emerges first and affects the other one, we designed two different architectures: *TBD-Social* (TBD-S) and *TBD-Task* (TBD-T). Both TBDs use a transfer learning approach to take advantage of the information learned beforehand on the dynamics of a specific dimension to predict the dynamics of the other one. Here, fItG is used as the pre-trained model. More specifically, TBD-S predicts the dynamics of social cohesion using a pre-trained fItG predicting the dynamics of task cohesion, whereas TBD-T predicts the dynamics of task cohesion using a pre-trained fItG prediciting the dynamics of social cohesion. TBD-T was already described in [13]. Figure 3a sketches the general TBD's architecture. It is composed of four components: *Input*, *Base*, *Target* and *Output* detailed in the following. Both TBD-S and TBD-T have a total of 49139 trainable weights.

Input is similar as in fItG: it is responsible for feature extraction. Features computed for each individual as well as for a group are distinctly processed by Base.

The Base component learns a representation of the group behavior for a dimension (i.e., social for TBD-T and Task for TBD-S) from which a group behavior representation for the targeted dimension (i.e., the predicted dimension) will be learned. For this purpose, Base uses a pre-trained version of the fItG model predicting the dynamics of only one dimension (i.e., social or task cohesion), and we enable the retraining of its weights. Base takes as input both individual and group features, and it outputs the representation of the group behavior learned for the specific dimension from the last layer of the Group component of the fItG model.

Target learns the group behavior representation of the targeted dimension (i.e., social or task cohesion). It consists of a FC layer with a ReLu activation function and 16 units that takes the output of Base as input. This FC layer is followed by five branches (one for each task). Each branch is composed of two consecutive FC layers with a ReLu activation function and eight and four units, respectively.

Finally, Output predicts the cohesion dynamics for the social dimension (in TBD-S) or the task dimension (in TBD-T), across the five tasks. It is composed of five branches (one for each task). Each branch consists of a FC layer with a sigmoid activation function and one unit, predicting the dynamics of one dimension for a specific task.

### 5.3 The TBD-RI architecture

Both TBDs integrate the social and task interplay unidirectionnally (i.e., from social to task cohesion with TBD-T and from Task to social cohesion with TBD-S). They, however, do not integrate the reciprocal impact of the two dimensions on each other. To try to integrate this reciprocity, we designed the *TBD-Reciprocal Impact* (TBD-RI) architecture. Built on top of both the TBD-S and TBD-T, TBD-RI also takes advantage of a transfer learning approach to learn a group behavior representation for each dimension before concatenating them and jointly learning the social and task cohesion dynamics. Figure 3b shows the TBD-RI architecture. It is composed of four components: *Input*, *Dimension Specific*, *Reciprocal Impact* and *Output*, and has 99002 trainable weights. Each component is detailed in the following.

As in the TBDs architectures, Input extracts the same features set to feed the two branches of the Dimension specific component.

Dimension Specific learns a representation of the group behavior for both social and task dimensions of cohesion. This component splits into two branches (i.e., one for each dimension) each one taking as input the same features extracted by Input. One branch learns the group behavior representation for task cohesion using a part of the TBD-T architecture, while the other branch learns the group behavior representation for social cohesion using a part of the TBD-S architecture. The parts of the TBDs that are used in the Dimension specific component are Base as well as the first FC layer of Target that contains 16 units. Then, the outputs from each branch are concatenated, resulting in a tensor of shape $[B \times T, 2 \times F]$ with $B$ the batch size (i.e., the number of group processed per batch), $T$ the number of timesteps (i.e., 6 timesteps per task, resulting in a total of 30 timesteps), and $F$ the size of the features representation of each dimension. This tensor is then processed by the Reciprocal impact component.

Reciprocal Impact learns the reciprocal impact that the social and task dimensions of cohesion have on each other over time using as input the concatenation of the representations learned by Dimension Specific. The component consists of a first FC layer with a ReLu activation function and 32 units, followed by another FC layer with a ReLu activation function of 16 units. Similar to the TBDs architecture,

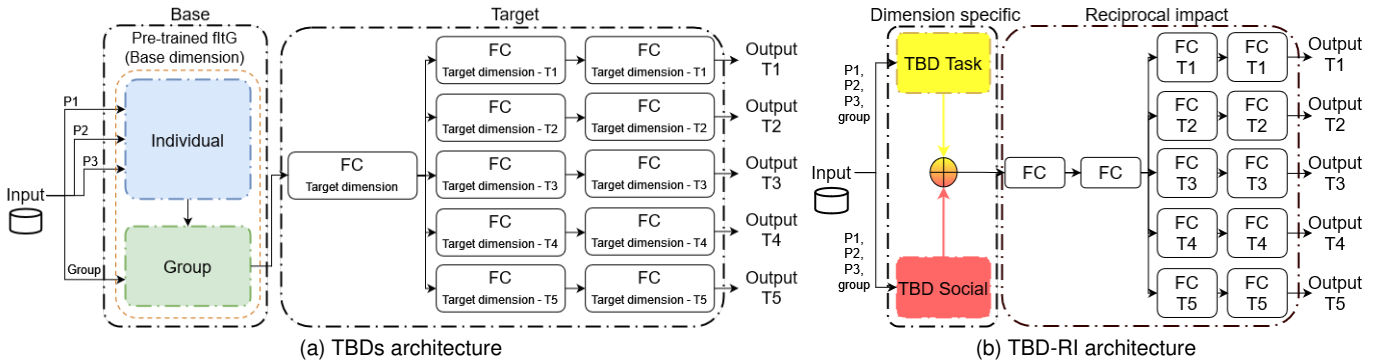



Fig. 3. The TBDs (a) and TBD-RI (b) architectures. Their Input component is similar to that one of fItG. (a) TBDs use, in Base, a pre-trained version of fItG dedicated to the prediction of the social (in TBD-T) or task (in TBD-S) cohesion dynamics by reusing its Individual (in blue) and Group (in green) components. Target learns a representation of the group behavior for the dynamics of the target dimension (i.e., social cohesion with TBD-S and task cohesion with TBD-Tor task cohesion) on top of the dynamics of the Base dimension, for each of the five tasks. (b) TBD-RI is built on top of TBD-S and TBD-T and learns, in Dimension Specific, a specific representation of the group behavior for each dimension. Both representations are concatenated and processed by Reciprocal Impact. As in fItG, Output predicts the social and task cohesion dynamics in a multilabel setting.

there is a split into five branches (one for each task) with two FC layers with a ReLu activation function and with eight and four units, respectively, in each branch.

Finally, similarly to fItG, Output consists of a FC layer, for each branch, with a sigmoid activation function and two units. This enables the RI to predict the dynamics of the social and task cohesion in a multilabel setting.

## 6 Evaluation

### 6.1 Methods

A Leave-One-Group-Out (LOGO) cross-validation was adopted. The reason is twofold: (i) the data size is small (15 groups), so we needed to guarantee to have enough training data for the learning, (ii) and we wanted to avoid biasing model performance by learning and testing with data coming from the same group. As reported in Section 4.1, every group is composed of three persons and the groups do not share persons. Thus, data was first split into training and test sets consisting of 14 and one group(s), respectively. Then, from the 14 groups of the training set, four groups were randomly picked and retained as validation set. The number of epochs was chosen, for each model, based on the highest average performance in predicting the cohesion dynamics on the groups of the validation set. To avoid overfitting and make the models more robust to noise, data augmentation was performed on the training set. Concretely, for each group, we added a Gaussian noise to all features ($\mu$= 0, $\sigma$= {0.01, 0.05}). We tested these two values of sigma to investigate the effect of such settings across the 15 rounds of the LOGO. For each seed, we retained the one that maximizes performance on the validation set[2], augmenting the data by a factor four. In addition, we also augmented the training set by computing all six permutations of the order of the group members to prevent models from learning undesirable patterns related to the order in which group members are processed. The final size of the training set was 240 groups. Building on preliminary studies, the models are trained up to a maximum of 500 epochs with a fixed learning rate of 0.001. The weights of the models are updated at every mini-batch. Each mini-batch is composed of four groups. Model performance is evaluated every 10 epochs on the validation set to determine the optimal number of epochs. Then, we evaluated model performance on the test set for each split of the LOGO (we recall that a test split includes no group/individual belonging to the training split).

We accounted for data imbalance by weighting a binary cross-entropy loss function in an inversely proportional way to the class frequencies as in Equation 3:

$$cw_{dim,T_x} = \frac{n_g}{n_c * n_{dim,T_x}} \quad (3)$$

where $n_g$ is the number of groups; $n_c$, the total number of classes (i.e., *decrease* and *not-decrease*) and $n_{dim,T_x}$, the number of occurrences for a class of dimension $dim$ in task $T_x$. The heuristic for computing the class weights in such a way is inspired by [79].

The models' performances were evaluated using F1 score as a metric. More specifically, in this work F1 is computed as the arithmetic mean of the per-class F1 scores. The F1 scores obtained from the 15 rounds of the LOGO cross-validation were averaged together giving an average F1 score value for each task. Moreover, the average F1 score across all tasks was also computed.

All the architectures presented in this study were developed and trained using Python 3.7 and Tensorflow 2.6 on NVIDIA V100 GPUs.

### 6.2 Comparing the models

As previously mentioned, the models' performances were evaluated over the 15 rounds of the LOGO cross-validation. To limit the randomness present in the models (e.g., due to the initialization of the weights and biases, and in regularization like dropouts), we followed recommendations from Colas and colleagues [80] that suggest evaluating models on several random seeds (between five and 25 depending on the data and algorithms) to obtain a reliable assessment of the models' performances. Here, we used 15 seeds and we computed the average and the standard deviation of the performance measures.

2. These values were chosen running preliminary studies.

We assessed potential significant differences between the performances through a computationally-intensive randomization test. This is a non-parametric test avoiding the independence assumption between the results being compared and that is suitable for non-linear measures such as F1-score [81]. We performed a k-sample permutation test using the *perm* package developed in R [82]. Such a test performs exact calculations using the Monte Carlo method during the permutation test. The significance level $\alpha$ was equal to 0.05. In case of multiple comparisons (i.e. comparisons between three models or between the five tasks), a posthoc analysis was carried out using pairwise permutation with a False Discovery Rate (FDR) adjusted p-value [83]. Such a p-value correction controls the false discovery rate, i.e., the expected proportion of false discoveries among the rejected hypotheses, hence integrating the rate of Type-I errors in the p-value computation.

## 7 Results

The overall average performances of TBDs and TBD-RI were first compared against those ones of fItG [14]. Then, a detailed analysis of the models' performances for the prediction of the dynamics of task and social cohesion during each task was carried out. Table 2 summarizes the overall average performance for each model as well as the performances for predicting the dynamics of each dimension of cohesion during task. Figure 4 shows the box-plots of the tasks' dynamics performances over the 15 seeds, for each model.

### 7.1 The dynamics of social cohesion

Regarding the dynamics of social cohesion, fItG reaches, on average over the 15 seeds, a F1-score of 0.67 $\pm$0.03, while TBD-S obtains 0.66 $\pm$0.04 and TBD-RI achieves 0.70 $\pm$0.03. A permutation test shows that there are significant differences in performance between these three models ($p = .018$). A posthoc analysis reveals that TBD-RI significantly outperformed both TBD-S ($p = .012$) and fItG ($p = .036$). Such an improvement is explained by the significant improvement on the dynamics prediction of T2 ($p = .044$), that is the only task in which TBD-RI significantly outperforms fItG (from 0.51 $\pm$0.13 to 0.61 $\pm$0.08). The dynamics of social cohesion in this task remains, however, the hardest to predict for all models. A permutation test run across the prediction of the five tasks shows a significant difference between the tasks performances for each model ($p = .001$ for every model). No significant differences are found between the predictions of dynamics in T1 and T2 across the models. These two tasks are, indeed, the ones for which all models obtained the lowest dynamics prediction performances. Models significantly perform better on T3 than on T1 ($p = .003$, for every model) and on T3 than on T2 ($p = .004$, $p = .003$ and $p = .013$ for the fItG, TBD-S, and TBD-RI, respectively). T4 and T5 are the tasks in which all models perform better when predicting the dynamics of social cohesion. ($p = .003$ between T3-T4 and $p = .003$ between T3-T5, for every model).

To summarize, TBD-RI is the best model for predicting the dynamics of social cohesion. TBD-RI significantly outperforms fItG and TBD-S, especially on T2. We observe a similar pattern on the performances obtained on each task across all models: T1 and T2 are the tasks for which the lowest performances are obtained in predicting the social dynamics, whatever the model; the dynamics of social cohesion in T3 is better predicted than in T1 and T2, while performances obtained in T4 and T5 are the highest ones.

### 7.2 The dynamics of task cohesion

Concerning the dynamics of task cohesion, a permutation test shows a significant difference in performance ($p = .014$) between fItG (0.64 $\pm$0.02 F1-score), TBD-T (0.66 $\pm$0.02 F1-score) and TBD-RI (0.64 $\pm$0.03 F1-score). A posthoc analysis reveals that the performance obtained by TBD-T is significant only with respect to that one of fItG ($p = .018$) but not with respect to that one of TBD-RI. Similarly to our findings regarding the dynamics of social cohesion, only one of the tasks for which the dynamics are the worst predicted is significantly improved as opposed to the baseline. In fact, TBD-T significantly outperforms fItG on predicting the dynamics in T3 ($p = .034$). TBD-T, indeed, reaches a F1-score equal to 0.69 $\pm$0.10 on T3 compared to the F1-score of 0.57 $\pm$0.13 obtained by fItG. Such improvement indicates a change in the ability of the models to predict the dynamics in a subset of tasks. Statistical analysis carried out through a permutation test shows a significant difference between the five tasks of every model ($p = .001$) for the fItG, TBD-T and TBD-RI, respectively. A posthoc analysis shows that, dynamics of task cohesion in T2 are significantly worst predicted than the ones in T4 ($p = .024$) and in T5 ($p = .007$) for the fItG, significantly worst predicted than in T3 ($p = .025$) and in T5 ($p = .010$) for TBD-T, and significantly worst predicted than in T5 for TBD-RI ($p = .005$). Dynamics in T5 remain significantly better-predicted across all the tasks and models (except than the ones in T3 in TBD-T). TBD-RI obtained less variations in performance across the tasks. There is only a significant difference between T5 and the other tasks ($p = .005$ for each pair of tasks T1-T5, T2-T5, T3-T5 and T4-T5), meaning that performances on T1, T2, T3, and T4 are equivalent.

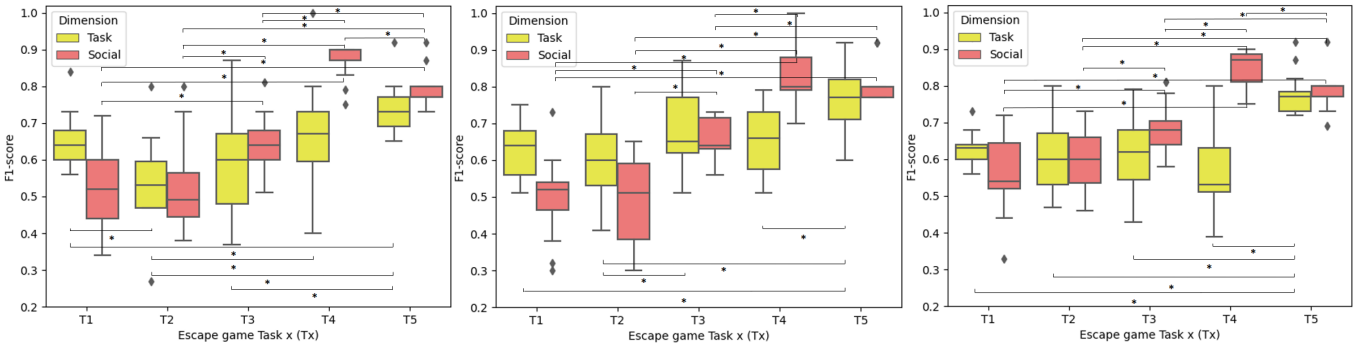
*To summarize, only TBD-T outperforms fItG, especially due to the significant improvement on T3. Also, dynamics of task cohesion in T2 remains among the worst predicted across all the models, while T5 is the task for which models significantly perform better according to F1-scores.*

## 8 Discussion

Since theoretical models of cohesion indicate that the social and task dimensions of cohesion are entangled, we conceived three architectures, TBD-S, TBD-T and TBD-RI, using transfer learning to jointly predict the dynamics of social and task cohesion. Transfer learning helps the training of the models in two ways. First, it enables them to start with a better weights' initialization, hence speeding up the training process. Then, it leverages knowledge from the pre-trained model on the dynamics of a specific dimension to predict the other one. The fact that, for the prediction of the dynamics of the task dimension, TBD-T outperformed fItG, illustrates this point and confirms insights from Social Science regarding stages of group development (i.e., social cohesion sets the stage for task cohesion). There is, however,

TABLE 2
Average F1-scores and standard deviations on the 15 seeds for each model, each task, and each dimension. The number between brackets next to each task is the rank of the performance obtained by the model over the five tasks. An equal number indicates that there is not a statistical difference between the performances. In bold the average performances that are significantly better across all the models.

| | Average F1-scores ±std (rank) | | | | | |
|---|---|---|---|---|---|---|
| | **fItG** | | **TBD-S/T** | | **TBD-RI** | |
| | **Social** | **Task** | **Social** | **Task** | **Social** | **Task** |
| **T1** | 0.52 ±0.10 (4) | 0.65 ±0.07 (2) | 0.50 ±0.11 (3) | 0.63 ±0.07 (2) | 0.56 ±0.10 (4) | 0.64 ±0.09 (2) |
| **T2** | 0.51 ±0.13 (4) | 0.56 ±0.12 (3) | 0.49 ±0.11 (3) | 0.59 ±0.09 (2) | 0.61 ±0.08 (4) | 0.61 ±0.09 (2) |
| **T3** | 0.65 ±0.07 (3) | 0.57 ±0.13 (3) | 0.66 ±0.06 (2) | 0.69 ±0.10 (1) | 0.69 ±0.06 (3) | 0.62 ±0.11 (2) |
| **T4** | 0.87 ±0.04 (1) | 0.66 ±0.14 (2) | 0.83 ±0.09 (1) | 0.65 ±0.09 (2) | 0.85 ±0.05 (1) | 0.57 ±0.10 (2) |
| **T5** | 0.80 ±0.04 (2) | 0.74 ±0.07 (1) | 0.80 ±0.05 (1) | 0.76 ±0.09 (1) | 0.79 ±0.05(2) | 0.78 ±0.05 (1) |
| **Average** | 0.67 ±0.03 | 0.64 ±0.02 | 0.66 ±0.04 | **0.66 ±0.02** | **0.70 ±0.03** | 0.64 ±0.03 |




(a) fItG (b) TBDs (c) TBD-RI

Fig. 4. Box-plots of the tasks' performances over the 15 seeds for fItG (4a), TBD-T and TBD-s (4b) and TBD-RI (4c). Significant differences between the prediction of dynamics of the tasks are marked with a "*". There is a similar pattern regarding the performances for predicting the dynamics of each task for the social dimension of cohesion: models obtain the worst performances in T1 and T2, while they reach the best performances in T4 and T5. Regarding the performances for predicting the dynamics of task cohesion, T2 remains one of the tasks for which the dynamics are the worst predicted across all the models, and T5 is always the one for which they are the best predicted.

no significant difference regarding the performances for predicting the dynamics of the social dimension between fItG and TBD-S. This also corroborates findings from Social Science regarding this specific direction of influence (i.e., task cohesion helps establish social cohesion) since it particularly concerns newly formed groups (e.g., [7], [43]). All the groups from the GAME-ON dataset are, indeed, groups of friends, limiting the applicability of such a theory. These results imply that task aspects, alone, cannot improve predictions of social cohesion because their relevance for understanding social cohesion hinges on their interdependence with social aspects. For example, when groups experience enjoyment due to their shared task focus, this can spill over into social cohesion. A shared task focus, without the mutually shared enjoyment of that experience (i.e. without the relational component), will, however, not affect social cohesion. Therefore, predicting the dynamics of cohesion benefits from integrating such an interplay with shared information. Depending on the task and the context, social or task aspects of cohesion might be predominant; hence, information from both dimensions is essential to improve predictions of the cohesion's dynamics and generalize to various tasks. The latter point is, indeed, confirmed by the TBD-RI performances for the dynamics of the social dimension. Such a model outperforms TBD-S by incorporating both TBD-S and TBD-T group behavior representations of each dimension and by combining them to learn the joint group behavior representation of both dimensions and predict the social and task dynamics of cohesion. These overall improvements of performances are mainly due to a significant improvement in performances on T2 by the TBD-RI regarding social cohesion and on T3 by TBD-T for task cohesion. Regarding T2, such an improvement might be explained by the task- and performance-driven nature of the task. As mentioned in Section 4, in this task, there was almost no social interaction among group members but they could see how successful each group member was. The group had to dispatch a set of enigmas to everyone and resolve a maximum of them on their own, without helping each other and they had to walk to a specific location to indicate that one enigma was solved. In this particular task, the sense of unity and the feeling of cohesion probably relied on the group task performances (i.e., how many enigmas each group member solved). In such a limited interaction, task cohesion provides essential information for predicting the dynamics of social cohesion since it is predominant. Regarding the improvement in performance on T3, we observe the opposite pattern. In this task, participants had to solve complex and contradictory enigmas under time pressure while helping each other. Such a setting challenged the members of the group in terms of organization and skills. Facing these difficulties, the groups might refocus themselves on the social aspects of cohesion to accomplish the task. For this particular task, using the knowledge learned about social cohesion to predict the dynamics of task cohesion is particularly efficient and helps understand the context.

## 9 Conclusion

In this paper, we computationally investigated cohesion, a multidimensional group emergent state typically considered as an affective construct in the literature, and originating from group members' behavioral interactions. We specifically focused on the dynamics of social and task cohesion and investigated their interplay over time. In this work, we adopted the GAME-ON dataset conceived to study the dynamics of task and social cohesion in groups of friends performing a variety of social activities organized in an escape game.

As discussed in the Social Science literature, a plethora of factors may shape the ways in which the social and the task dimensions of cohesion are interrelated. In some cases, social cohesion may be a driver of task cohesion, especially in groups with strong social bonds, while the contrary might hold true for more task-focused groups (cf. [22], [44]). To account for these different possibilities of mutual influence between the task and social dimensions of cohesion, we developed three DNNs: TBD-S, TBD-T, and TBD-RI. They adopt a transfer learning approach to take advantage of the information learned for one dimension to predict the dynamics of the other one. TBD-RI, exploiting both TBD-S and TBD-T, offers a way to integrate the reciprocal impact of both dimensions on each other. These DNNs were evaluated against a state-of-the-art model that predicts the dynamics of social and task cohesion in a multilabel setting. Results show that, for the prediction of the dynamics of the social dimension, TBD-RI outperformed the other models, especially on T2, while for the task dimension, TBD-T achieved significantly better overall performance than the baseline, in particular on T3.

While the obtained results corroborate insights from Social Science, our work also has some limitations. First, our sample comprised groups of friends. Further investigation is needed to understand to what extent our results apply to different types of groups. For example, in groups of work colleagues, task cohesion might emerge before social cohesion. Moreover, our architectures are designed to process groups of a specific size (here three persons), which limits the applicability of our approach to the analysis of cohesion in larger groups or with a number of members changing over time. Also, our architectures focused on predicting social and task cohesion dynamics on a set of tasks that did not require particular skills and in which group members were freely interacting, hence, reflecting the activities that social groups might encounter. Further analysis is needed to evaluate if our architectures adapt to more complex tasks in other environments where group members movements and actions are limited (e.g., during a meeting). Finally, cohesion is a multidimensional emergent state with additional facets beyond the most frequently investigated task and social dimensions, which we also focused on here. Group pride, for example, is another cohesion dimension that could be investigated from a computational point of view as well as the way it could interrelated with the social and the task dimensions for developing more exhaustive computational models of cohesion.

## Acknowledgments

This work was 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] M. J. Waller, G. A. Okhuysen, and M. Saghafian, "Conceptualizing emergent states: A strategy to advance the study of group dynamics," *The Academy of Management Annals*, vol. 10, no. 1, pp. 561–598, 2016.

[2] M. A. Marks, J. E. Mathieu, and S. J. Zaccaro, "A temporally based framework and taxonomy of team processes," *The Academy of Management Review*, vol. 26, no. 3, pp. 356–376, 2001.

[3] S. W. Kozlowski and D. R. Ilgen, "Enhancing the effectiveness of work groups and teams," *Psychological science in the public interest*, vol. 7, no. 3, pp. 77–124, 2006.

[4] R. Grossman, S. B. Friedman, and S. Kalra, "Teamwork processes and emergent states," *The wiley blackwell handbook of the psychology of team working and collaborative processes*, vol. 42, pp. 243–269, 2017.

[5] S. W. J. Kozlowski and G. T. Chao, "The dynamics of emergence: Cognition and cohesion in work teams," *Managerial and Decision Economics*, vol. 33, no. 5-6, pp. 335–354, 2012.

[6] E. Salas, R. Grossman, A. M. Hughes, and C. W. Coultas, "Measuring team cohesion: Observations from the science," *Human Factors*, vol. 57, no. 3, pp. 365–374, 2015.

[7] A. V. Carron and L. R. Brawley, "Cohesion: Conceptual and measurement issues," *Small Group Research*, vol. 31, pp. 89–106, 2000.

[8] S. M. Fiore, D. R. Carter, and R. Asencio, "Conflict, trust, and cohesion: Examining affective and attitudinal factors in science teams," in *Team cohesion: Advances in psychological theory, methods and practice*. Emerald Group Publishing Limited, 2015.

[9] T. Rapp, T. Maynard, M. Domingo, and E. Klock, "Team emergent states: What has emerged in the literature over 20 years," *Small Group Research*, vol. 52, no. 1, pp. 68–102, 2021.

[10] J. E. Mathieu, J. R. Hollenbeck, D. van Knippenberg, and D. R. Ilgen, "A century of work teams in the journal of applied psychology." *Journal of applied psychology*, vol. 102, no. 3, p. 452, 2017.

[11] N. Lehmann-Willenbrock and H. Hung, "A multimodal social signal processing approach to team interactions," *Organizational Research Methods*, p. 10944281231202741, 2023.

[12] L. Maman, E. Ceccaldi, N. Lehmann-Willenbrock, L. Likforman-Sulem, M. Chetouani, G. Volpe, and G. Varni, "Game-on: A multimodal dataset for cohesion and group analysis," *IEEE Access*, vol. 8, pp. 124 185–124 203, 2020.

[13] L. Maman, L. Likforman-Sulem, M. Chetouani, and G. Varni, "Exploiting the interplay between social and task dimensions of cohesion to predict its dynamics leveraging social sciences," in *Proceedings of the 23rd International Conference on Multimodal Interaction*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 16–24.

[14] L. Maman, M. Chetouani, L. Likforman-Sulem, and G. Varni, "Using valence emotion to predict group cohesion's dynamics: Top-down and bottom-up approaches," in *Proceedings of the 9th International Conference on Affective Computing and Intelligent Interaction*, 2021, pp. 1–8.

[15] K. Lewin, "Field theory and experiment in social psychology: Concepts and methods," *American Journal of Sociology*, vol. 44, no. 6, pp. 868–896, 1939.

[16] K. W. Back, "Influence through social communication," *The Journal of Abnormal and Social Psychology*, vol. 46, no. 1, pp. 9–23, 1951.

[17] D. J. Beal, R. R. Cohen, M. J. Burke, and C. L. McLendon, "Cohesion and performance in groups: A meta-analytic clarification of construct relations," *Journal of Applied Psychology*, vol. 88, no. 6, pp. 989–1004, 2003.

[18] K. L. Dion, "Group cohesion: From field of forces to multidimensional construct," *Group Dynamics: Theory, Research, and Practice*, vol. 4, no. 1, pp. 7–26, 2000.

[19] M. T. Maynard, D. M. Kennedy, S. A. Sommer, and A. M. Passos, "Team cohesion: A theoretical consideration of its reciprocal relationships within the team adaptation nomological network," in *Team cohesion: Advances in psychological theory, methods and practice*. Emerald Group Publishing Limited, 2015, vol. 17, pp. 83–111.

[20] J. Griffith, "Measurement of group cohesion in u.s. army units," *Basic and Applied Social Psychology*, vol. 9, no. 2, pp. 149–171, 1988.

[21] J. Griffith and M. Vaitkus, "Relating cohesion to stress, strain, disintegration, and performance: An organizing framework," *Military Psychology*, vol. 11, no. 1, pp. 27–55, 1999.

[22] J. B. Severt and A. X. Estrada, "On the function and structure of group cohesion," in *Team Cohesion: Advances in Psychological Theory, Methods and Practice*. Emerald Group Publishing Limited, 2015, vol. 17, pp. 3–24.

[23] K. A. Bollen and R. H. Hoyle, "Perceived cohesion: A conceptual and empirical examination," *Social Forces*, vol. 69, no. 2, pp. 479–504, 1990.

[24] J. Keyton, "Relational communication in groups," *The handbook of group communication theory and research*, pp. 192–222, 1999.

[25] J. Keyton, S. J. Beck, M. S. Poole, and D. S. Gouran, "Group communication: A continued evolution," in *The Emerald Handbook of Group and Team Communication Research*. Emerald Publishing Limited, 2021.

[26] N. C. Magpili and P. Pazos, "Self-managing team performance: A systematic review of multilevel input factors," *Small Group Research*, vol. 49, no. 1, pp. 3–33, 2018.

[27] H. Hung and D. Gatica-Perez, "Estimating cohesion in small groups using audio-visual nonverbal behavior," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 563–575, 2010.

[28] M. C. Nanninga, Y. Zhang, N. Lehmann-Willenbrock, Z. Szlávik, and H. Hung, "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, 2017, pp. 206–215.

[29] M. Stel and R. Vonk, "Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction," *British journal of psychology*, vol. 101, no. 2, pp. 311–323, 2010.

[30] A. Dhall, "Emotiw 2019: Automatic emotion, engagement and cohesion prediction tasks," in *Proceedings of the 21st International Conference on Multimodal Interaction*, 2019, pp. 546–550.

[31] D. Guo, K. Wang, J. Yang, K. Zhang, X. Peng, and Y. Qiao, "Exploring regularizations with face, body and image cues for group cohesion prediction," in *Proceedings of the 21st International Conference on Multimodal Interaction*. Association for Computing Machinery, 2019, pp. 557–561.

[32] B. Zou, Z. Lin, H. Wang, Y. Wang, X. Lyu, and H. Xie, "Joint prediction of group-level emotion and cohesiveness with multi-task loss," in *Proceedings of the 5th International Conference on Mathematics and Artificial Intelligence*, 2020, pp. 24–28.

[33] G. Sharma, S. Ghosh, and A. Dhall, "Automatic group level affect and cohesion prediction in videos," in *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos*. IEEE, 2019, pp. 161–167.

[34] S. Ghosh, A. Dhall, N. Sebe, and T. Gedeon, "Automatic prediction of group cohesiveness in images," *IEEE Transactions on Affective Computing*, pp. 1–1, 2020.

[35] S. G. Barsade and D. E. Gibson, "Group emotion: A view from top and bottom." *D. H Gruenfeld (Ed.)*, vol. 1, pp. 81–102, 1998.

[36] B. W. Tuckman, "Developmental sequence in small groups," *Psychological bulletin*, vol. 63, no. 6, pp. 384–399, 1965.

[37] R. Grossman, Z. Rosch, D. Mazer, and E. Salas, "What matters for team cohesion measurement? A Synthesis," *Research on Managing Groups and Teams*, vol. 17, pp. 147–180, 2015.

[38] S. W. Kozlowski, S. M. Gully, E. R. Nason, and E. M. Smith, "Developing adaptive teams: A theory of compilation and performance across levels and time," *Pulakos (Eds.), The Changing Nature of Performance: Implications for Staffing, Motivation, and Development*, pp. 240–292, 1999.

[39] P. J. Runkel, M. Lawrence, S. Oldfield, M. Rider, and C. Clark, "Stages of group development: An empirical test of tuckman's hypothesis," *The Journal of Applied Behavioral Science*, vol. 7, no. 2, pp. 180–193, 1971.

[40] L. A. Zurcher Jr, "Stages of development in poverty program neighborhood action committees," *The Journal of Applied Behavioral Science*, vol. 5, no. 2, pp. 223–258, 1969.

[41] B. W. Tuckman and M. A. C. Jensen, "Stages of small-group development revisited," *Group & Organization Studies*, vol. 2, no. 4, pp. 419–427, 1977.

[42] I. Balaguer, I. Castillo, and J. Duda, "Interrelationships between motivational climate and cohesion in cadet football," *EduPsykhe*, vol. 2, no. 2, pp. 243–58, 2003.

[43] M. Boyd, M.-S. Kim, N. Ensari, and Z. Yin, "Perceived motivational team climate in relation to task and social cohesion among male college athletes," *Journal of Applied Social Psychology*, vol. 44, no. 2, pp. 115–123, 2014.

[44] A. V. Carron, W. N. Widmeyer, and L. R. Brawley, "The development of an instrument to assess cohesion in sport teams: The group environment questionnaire," *Journal of Sport Psychology*, vol. 7, no. 3, pp. 244–266, 1985.

[45] M. A. Eys, J. Hardy, A. V. Carron, and M. R. Beauchamp, "The relationship between task cohesion and competitive state anxiety," *Journal of Sport and Exercise Psychology*, vol. 25, no. 1, pp. 66–76, 2003.

[46] S. A. Kozub and J. F. McDonnell, "Exploring the relationship between cohesion and collective efficacy in rugby teams," *Journal of sport behavior*, vol. 23, no. 2, pp. 120–129, 2000.

[47] U. Kubasova, G. Murray, and M. Braley, "Analyzing verbal and nonverbal features for predicting group performance," *arXiv preprint arXiv:1907.01369*, 2019.

[48] A. M. Alsulami, "Towards building group cohesion and learning outcomes based on nonverbal immediacy behavior," *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, vol. 12, no. 7, pp. 1–12, 2021.

[49] S. J. Kamper, R. W. Ostelo, D. L. Knol, C. G. Maher, H. C. de Vet, and M. J. Hancock, "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*, vol. 63, no. 7, pp. 760–766, 2010.

[50] R. G. Lord, J. F. Binning, M. C. Rush, and J. C. Thomas, "The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior," *Organizational Behavior and Human Performance*, vol. 21, no. 1, pp. 27–39, 1978.

[51] D. Gatica-Perez, I. McCowan, D. Zhang, and S. Bengio, "Detecting group interest-level in meetings," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 2005, pp. 489–492.

[52] E. Ceccaldi, N. Lehmann-Willenbrock, E. Volta, M. Chetouani, G. Volpe, and G. Varni, "How unitizing affects annotation of cohesion," in *Proceedings of the 8th International Conference on Affective Computing and Intelligent Interaction*. IEEE, 2019, pp. 1–7.

[53] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.

[54] E. T. Hall, *The hidden dimension*. Anchor, 1966, vol. 609.

[55] A. Hans and E. Hans, "Kinesics, haptics and proxemics: Aspects of non-verbal communication," *IOSR Journal of Humanities and Social Science*, vol. 20, no. 2, pp. 47–52, 2015.

[56] N. L. Ashton, M. E. Shaw, and A. P. Worsham, "Affective reactions to interpersonal distances by friends and strangers," *Bulletin of the Psychonomic Society*, vol. 15, no. 5, pp. 306–308, 1980.

[57] A. Kendon, "Spatial organization in social encounters: The f-formation system," *Conducting interaction: Patterns of behavior in focused encounters*, pp. 209–238, 1990.

[58] R. L. Birdwhistell, *Kinesics and context*. University of Pennsylvania press, 2010.

[59] S. Goldin-Meadow and M. W. Alibali, "Gesture's role in speaking, learning, and creating language," *Annual review of psychology*, vol. 64, pp. 257–283, 2013.

[60] C. Carmeli, M. G. Knyazeva, G. M. Innocenti, and O. De Feo, "Assessment of eeg synchronization based on state-space analysis," *Neuroimage*, vol. 25, no. 2, pp. 339–354, 2005.

[61] H. G. Wallbott, "Bodily expression of emotion," *European journal of social psychology*, vol. 28, no. 6, pp. 879–896, 1998.

[62] S. Piana, M. Mancini, A. Camurri, G. Varni, and G. Volpe, "Automated analysis of non-verbal expressive gesture," in *Human Aspects in Ambient Intelligence*. Springer, 2013, pp. 41–54.

[63] G. E. Weisfeld and J. M. Beresford, "Erectness of posture as an indicator of dominance or success in humans," *Motivation and Emotion*, vol. 6, no. 2, pp. 113–131, 1982.

[64] J. L. Tracy and R. W. Robins, "Show your pride: Evidence for a discrete emotion expression," *Psychological Science*, vol. 15, no. 3, pp. 194–197, 2004.

[65] J. Aigrain, M. Spodenkiewicz, S. Dubuisson, M. Detyniecki, D. Cohen, and M. Chetouani, "Multimodal stress detection from multi-

ple assessments," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 491–506, 2018.
[66] A. Saarinen, V. Harjunen, I. Jasinskaja-Lahti, I. P. Jääskeläinen, and N. Ravaja, "Social touch experience in different contexts: A review," *Neuroscience & Biobehavioral Reviews*, vol. 131, pp. 360–372, 2021.
[67] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
[68] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.
[69] F. Ringeval, E. Marchi, C. Grossard, J. Xavier, M. Chetouani, D. Cohen, and B. Schuller, "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, 2016, pp. 1210–1214.
[70] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, "Long short term memory recurrent neural network based multimodal dimensional emotion recognition," in *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*. New York, NY, USA: Association for Computing Machinery, 2015, pp. 65—72.
[71] M. Goudbeek and K. Scherer, "Beyond arousal: Valence and potency/control cues in the vocal expression of emotion," *The Journal of the Acoustical Society of America*, vol. 128, no. 3, pp. 1322–1336, 2010.
[72] F. Weninger, F. Eyben, B. W. Schuller, M. Mortillaro, and K. R. Scherer, "On the acoustics of emotion in audio: what speech, music, and sound have in common," *Frontiers in psychology*, vol. 4, pp. 1–12, 2013.
[73] M. Tahon and L. Devillers, "Acoustic measures characterizing anger across corpora collected in artificial or natural context," in *Speech Prosody*, 2010.
[74] F. Eyben, F. Weninger, F. Gross, and B. Schuller, "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 835–838.
[75] K. Hilton, "The perception of overlapping speech: Effects of speaker prosody and listener attitudes," in *Proceedings of Interspeech*, 2016, pp. 1260–1264.
[76] K. Ryokai, E. Duran Lopez, N. Howell, J. Gillick, and D. Bamman, "Capturing, representing, and interacting with laughter," in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–12.
[77] R. R. Provine, "Laughter punctuates speech: Linguistic, social and gender contexts of laughter," *Ethology*, vol. 95, no. 4, pp. 291–298, 1993.
[78] P. Glenn, *Laughter in interaction*. Cambridge University Press, 2003, vol. 18.
[79] G. King and L. Zeng, "Logistic regression in rare events data," *Political analysis*, vol. 9, no. 2, pp. 137–163, 2001.
[80] C. Colas, O. Sigaud, and P.-Y. Oudeyer, "How many random seeds? statistical power analysis in deep reinforcement learning experiments," 2018.
[81] A. Yeh, "More accurate tests for the statistical significance of result differences," *The 18th International Conference on Computational Linguistics*, vol. 2, pp. 947–954, 2000.
[82] M. P. Fay and P. A. Shaw, "Exact and asymptotic weighted logrank tests for interval censored data: The interval r package," *Journal of Statistical Software*, vol. 36, no. 2, pp. 1–34, 2010.
[83] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

**Lucien MAMAN** received both his MSc degree in Software Engineering for Technical Computing from Cranfield University, U.K., and his engineering diploma from ESTIA, France in 2017. After working for two years as a video engineer at Grabyo, U.K., he completed his Ph.D. in 2022 at the LTCI, Télécom Paris, Institut Polytechnique de Paris, France. His research interests include Social Signal Processing, Machine Learning, Deep Learning, Affective Computing, Emergent States, and Cohesion.

**Nale LEHMANN-WILLENBROCK** is a Full Professor and department head of Industrial/Organizational Psychology and leads the TeamLab at the University of Hamburg, Germany. Previously, she was an Associate Professor at the University of Amsterdam, The Netherlands. She holds a Ph.D. in Psychology from Technical University Braunschweig (2012). She investigates team processes and leader-follower dynamics during organizational meetings and other interaction settings using pattern analytical methods and promotes interdisciplinary collaborations that bridge social and computer science. She currently serves as Associate Editor at Small Group Research.

**Mohamed CHETOUANI** is currently a Full Professor in signal processing and machine learning for human-machine interaction. He is affiliated to the PIRoS (Perception, Interaction et Robotique Sociales) research team at the Institute for Intelligent Systems and Robotics (CNRS UMR 7222), Sorbonne University (formerly Pierre and Marie Curie University). His activities cover social signal processing, social robotics and interactive machine learning with applications in psychiatry, psychology, social neuroscience and education. Since 2018, he is the coordinator of the ANIMATAS H2020 Marie Sklodowska Curie European Training Network. Since 2019, he is the President of Sorbonne University Ethical Committee. He was involved in several educational activities including organization of summer schools. He is member of the management board of the International AI Doctoral Academy initiated by European networks of AI excellence centers. He is member of the EU Network of Human-Centered AI. He was Program co-chair of ACM ICMI 2020. He was General Chair of ACM ICMI 2023. He serves as Associate Editor of IEEE Transactions on Affective Computing.

**Laurence LIKFORMAN-SULEM** is graduated in engineering from ENST-Bretagne (Ecole Nationale Superieure des Telecommunications) in 1984, received her Ph.D. from ENST-Paris in 1989, and her HDR (Habilitation a Diriger des Recherches) from Pierre & Marie Curie University in 2008. She is an Associate Professor at Telecom ParisTech in the IDS (Image Data Signal) Department since 1991 where she serves as a senior instructor in Pattern Recognition and handwriting recognition. She chaired the program committee of CIFED held in Fribourg, Switzerland, in 2006, the program committees of two DRR Conferences (Document Recognition and Retrieval) held in 2009 and 2010 in San Jose, California, and the ASAR (Arabic and derived Script Analysis and Recognition) workshop held in 2017 in Nancy, France.

**Giovanna VARNI** is an Associate Professor at Department of Information Engineering and Computer Science (DISI), University of Trento, Italy. Previously she was an Associate Professor at LTCI, Télécom Paris, Institut polytechnique de Paris, France. She is an interdisciplinary researcher mainly investigating on social Signal Processing (SSP) and Human Computer Interaction (HCI). She was involved in several EU FP7-FP6 projects and she was PI of the national French project ANR JCJC GRACE (2019-2022) on the automated analysis of cohesion in small groups of humans. She contributes regularly to organizational roles in international conferences and workshops relevant for her specific research area such as ACII and ICMI, for which she also serves as a Program Committee member.