Artificial Intelligence and Social Computing, Vol. 122, 2024, 103–113
https://doi.org/10.54941/ahfe1004647

Dyadic Interactions and Interpersonal Perception: An Exploration of Behavioral Cues for Technology-Assisted Mediation

Hifza Javed[1], Nina Moorman[1,2], Thomas Weisswange[3], and Nawid Jamali[1]

[1]Honda Research Institute USA, Inc, San Jose, CA 95134, USA
[2]Georgia Institute of Technology, Atlanta, GA 30332, USA
[3]Honda Research Institute Europe GmbH, 63073 Offenbach am Main, Germany

ABSTRACT

Mediators aim to shape group dynamics in various ways, such as improving trust and cohesion, balancing participation, and promoting constructive conflict resolution. Technological systems used to mediate human-human interactions must be able to continuously assess the state of the interaction and generate appropriate actions. In this paper, behavioral cues that indicate interpersonal perception in dyadic social interactions are investigated. These cues may be used by such systems to produce effective mediation strategies. These are used to evaluate dyadic interactions, in which each interactant rates how agreeable or disagreeable the other interactant comes across. A multi-perspective approach is taken to evaluate interpersonal affect in dyadic interactions, employing computational models to investigate behavioral cues that reflect interpersonal perception in both the interactant providing the rating and the interactant being rated. The findings offer nuanced insights into interpersonal dynamics, which will be beneficial for future work on technology-assisted social mediation.

Keywords: Affect, Interpersonal perception, Group interaction, Social mediation, Feature importance, Behavioral cues

INTRODUCTION

Social interaction with other humans constitutes a significant part of our lives. How positively such interactions are perceived is influenced by complex inter-personal dynamics, context, conversation topics, mood, and many other factors (Forsyth, 2014). Emotional reactivity of social partners can also give rise to various dynamic interpersonal emotional patterns (Van Kleef & Côté, 2022). Mediation aims to shape group dynamics such as improving trust and cohesion, and balancing participation, to improve the overall satisfaction of all participants within a social interaction.

In recent years, researchers have been investigating the application of technology to facilitate social mediation. Understanding the collective affective state of a group is necessary for technology-assisted systems to improve group dynamics. Traditional methods for detecting emotion of individuals are unable to appropriately capture interpersonal dynamics in multiparty interactions (Jung, 2017). Previous approaches for group affect recognition predominantly estimate overall group states by utilizing annotations from external observers. These approaches determine individual affective states of group members and combine them to form a single metric for group affect (Veltmeijer, et al., 2021). However, for technology-assisted mediation to be efficient, assistance should be based on the judgment of the group members themselves and must capture interpersonal affective states rather than the internal affect of each group member.

This work focuses on estimating interpersonal affect within human-human interactions. It utilizes computational modeling to investigate behavioral cues that serve as indicators of a person's perception of their partner within dyadic interactions. Specifically, a dataset collected using the COntinuous Retrospective Affect Evaluation (CORAE) tool (Sack, et al., 2023) is utilized, which contains dyadic interaction data along with continuous labels of interpersonal affect. This dataset is unique in that it provides ratings of affect that, first, are provided by the interactants themselves rather than external observers, second, record interpersonal perception from the perspective of each individual, and third, evaluate their perception of their partner rather than their own affective states. This offers an opportunity to study how observable behavioral cues that reflect interpersonal affect in an interaction may vary between the person providing the rating and the person being rated. A study is conducted into the behavioral cues that inform the interpersonal perception and address the following research questions:

- RQ1: How do the behavioral cues that reflect interpersonal perception vary between an interactant's self-behavior and their partner's behavior?
- RQ2: Which behavioral cues are important indicators of interpersonal perception in self and partner behaviors?
- RQ3: Do the important behavioral cues within self and partner behaviors differ between the positively- and negatively-rated interactions?

The proposed multi-perspective approach analyzes interactions from the view of the self and the partner. The findings inform which self and partner features to leverage to determine interpersonal perception in order to produce effective mediation strategies. This work offers nuanced insights into interpersonal dynamics, beneficial for future work on technology-assisted social mediation.

METHODOLOGY

To gain insight into the behavioral cues that serve as indicators of interpersonal affect, a computational approach was adopted. The problem was formulated as a classification task to predict interpersonal affect ratings from audiovisual features.

Dataset and Feature Extraction

To answer the research questions, a dyadic interaction dataset, specifically designed for assessing affective responses in the context of interpersonal interactions (Sack, et al., 2023) was used. The dataset consists of virtual dyadic interactions (Figure 1 (left)), approximately 10 minutes long, between English-speaking adults as they engage in a collaborative decision-making task to rank various reasons for poverty (Shek, 2002). Using the CORAE tool, each interactant retrospectively rates how their partner came across during the interaction along a 15-point scale ranging from disagreeable (−7) to agreeable (+7). This perception of the interaction partner is referred to as interpersonal affect. This results in two ratings (one for each interactant) per timestamp for each session. Since both interactants' perspectives are accesisble at any given time, a multi-perspective approach consisting of the ego and the partner perspectives can be utilized. The dataset consists of 30 interaction sessions in total.

Multimodal behavioral features, including visual and audio features (Figure 1 (center)) were extracted. Using iMotions' facial expression module, 34 two-dimensional coordinates of facial landmarks were extracted (Appendix A). The data were processed at an average rate of 30 frames per second. Out of the available 30 sessions, 26 were used in the analysis. The remaining 4 were excluded due to difficulties in processing the video data for visual feature extraction.

For audio, short-term spectral features were extracted. Though both global and short-term acoustic features have been utilized in prior research, global-level acoustic features are limited in their ability to describe the short-term, dynamic variations (Busso, et al., 2004) that commonly occur within a human-human social interaction. Therefore, 34 audio features that represent low-level descriptors of voice were extracted, including Mel Frequency Cepstral Coefficients (MFCCs), energy, and spectral features (Appendix A). These features were extracted separately for each individual. Thereafter, feature normalization was performed for each interactant. The resulting dataset contained a total of 433606 samples.

Model

For the multi-class classification of interpersonal affect, the 15 discrete rating labels were first one-hot-encoded. Since the dataset analysis showed class imbalance, a random forest classifier was chosen as it is known to be effective for imbalanced data (Khoshgoftaar, et al., 2007). The random forest classifier available in the sci-kit learn library[1] was used, with the number of decision trees in the forest set to 100. Additionally, Gini impurity was employed as the criterion to assess the quality of splits in each tree. Figure 1 depicts the complete modeling pipeline.

Experiments

This section outlines a series of experiments designed to investigate the research questions. Proposing a specific modeling approach is beyond the scope of this work. Instead, it focuses on analyzing the dataset's predictive power for interpersonal affect classification and providing insights for future modeling efforts with this dataset.

[1] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

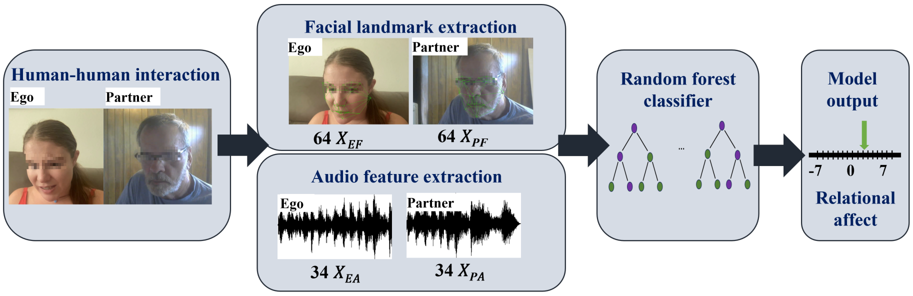

Figure 1: An illustration of the modelling approach starting with the dyadic human-human interactions (left), audiovisual feature extraction (center), and random forest-based classification of relational affect (right). Here, X_{EF}: ego facial features, X_{EA}: ego audio features, X_{PF}: partner facial features, and X_{PA}: partner audio features.

Ego, partner, and joint models of interpersonal affect. To study the importance of self-behavior and partner's behavior in predicting interpersonal affect, an experiment was designed to utilize three models for classifying interpersonal affect: 1) the ego model, 2) the partner model, and 3) the joint model. The ego model investigates how much of an interactant's self-behavior reflects their interpersonal perception of their partner. To achieve this, a model that uses ego features (X_E) to predict ratings from ego perspective (Y_E) is learned. Conversely, the partner model investigates how much of the partner behavior reflects the interpersonal ratings they received from the other interactant. To achieve this, a model that uses partner features (X_P) to predict ego ratings (Y_E) is learned. Additionally, a joint model is learned in order to investigate whether the combination of both interactants' behaviors can capture interpersonal dynamics that the individual models may not. The combined feature space may better capture the interplay of behaviors such as gaze, tone of voice, backchanneling, etc. that are inherent to interpersonal exchanges in dyadic interactions. To achieve this, a model that uses both interactants' features (X_E and X_P) to predict ego ratings (Y_E) is used. To ensure a robust model assessment, a 10-fold cross-validation with a 90%/10% test-train split is performed. Table 1 summarizes the features for each model.

Table 1. Features and labels used in the ego, partner, and joint models.

Model name	Feature space	Labels
Ego model	Ego features: 64 X_{EF}, 34 X_{EA}	Ego ratings, Y_E
Partner model	Partner features: 64 X_{PF}, 34 X_{PA}	Ego ratings, Y_E
Joint model	Ego and partner features: 64 X_{EF}, 34 X_{EA}, 64 X_{PF}, 34 X_{PA}	Ego ratings, Y_E

Feature Importance Analyses. The goal in this experiment was to identify key behavioral cues contributing to the prediction of interpersonal affect. To this end, feature importance analysis was conducted on each of the three models, yielding an importance score for every feature in the model's respective feature space. A higher score indicates a greater impact on the model's performance. Permutation-based feature importance analysis was used, where the importance of a feature is signified by the decrease in model accuracy when this feature value is randomly shuffled. This drop in the model's performance indicates how much the model depends on that feature. The results of this analysis enable a comparison of the features in both self and partner behaviors that are most reflective of the prevailing interpersonal affect within an interaction, as represented in the three models.

While the landmarks provide a granular view into the facial activity, it was anticipated that a higher-level analysis of the facial expressions may also be useful in contextualizing the findings from the raw landmarks. Therefore, facial action units were utilized, since these represent the fundamental muscular activity that produces facial appearance changes (Ekman, 1978). A total of 24 action units were extracted, including brow raise, chin raise, jaw drop, head pitch, yaw, and roll, among others (Appendix A). Feature importance analysis was then conducted to offer additional insights for the design of robust mediation systems.

Feature importance analyses for negative and positive interpersonal affect. The goal in this experiment was to identify specific behavioral cues that capture negative versus positive interpersonal affect within dyadic interactions. To achieve this, the dataset was split such that all samples with ratings between −1 and −7 become part of the positive dataset and all those containing ratings between +1 and +7 are assigned to the negative dataset. Feature importance analysis was then conducted on the two datasets using all three models. It is important to note that the sizes of the two data splits in this experiment are different, where the positive dataset consists of 349711 samples and the negative dataset consists of 18731 samples. The remaining data samples contained labels for neutral interpersonal affect (0 rating value). Findings from this experiment shed light on the differences in behavioral expression of positive and negative interpersonal affect in dyadic interactions, both from the ego and partner perspectives.

RESULTS

Ego, Partner, and Joint Models of Interpersonal Affect

This experiment aimed to investigate the importance of self-behavior and partner's behavior in predicting interpersonal affect (RQ1). Figure 2 shows that the random forest model was able to classify the interpersonal affect ratings with an overall accuracy of above 76% for all three modes. This indicates that the feature representations used in these experiments can capture the relevant behavioral cues needed for the evaluation of interpersonal affect in all three cases. The joint model was found to perform with a higher accuracy (86.5%) than either of the individual models (ego: 77.4% and partner: 76.8%). These results suggest that including both interactants' observations when predicting interpersonal affect effectively captures the interplay of behaviors between the interactants. Additionally, the findings suggest that the individual model can still be useful in cases where one interactant is occluded. Additionally, competitive performance between the ego and the partner models was found, with the ego model resulting in a higher average accuracy than the partner model. These findings suggest that behavioral evidence of interpersonal affect is more pronounced in the features of the interactant providing the ratings rather than the interactant being rated.

The Wilcoxon signed rank test was used to determine if the differences in performance of the three models were statistically significant. Since the same interactants are represented in the data subsets used in the three models, this test was deemed appropriate to compare these related data. It is found that all three model comparisons are significant, with p-values < 0.01.

Feature Importance Analysis

The aim of this experiment was to identify the key behavioral cues contributing to the prediction of interpersonal affect (RQ2). The results of this experiment provide insight into the predictive power of the features used in each model. First, the importance of ego and partner features was investigated. To this end, the relative importance scores of all ego and partner features were computed and the ego (48.8%) and partner (51.2%) feature importance scores were found to be comparable. This implies that both interactants' behavioral features contribute similarly to the model's performance in predicting interpersonal affect in this model.

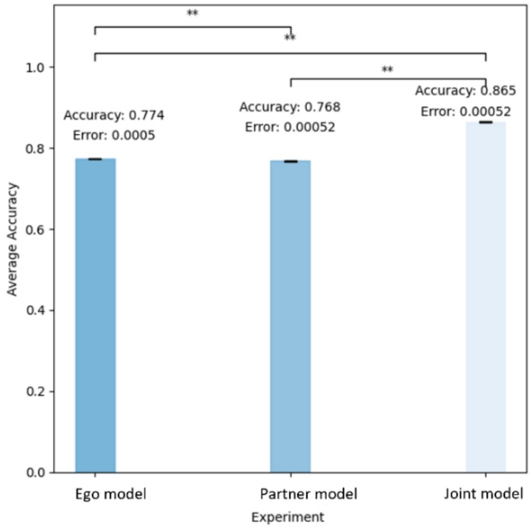

Figure 2: Comparison of accuracies of the ego, partner, and joint models. Statistically significant differences are found between the three models using the Wilcoxon signed rank test, where ** represents a p-value < 0.01.

Next, the importance of facial and audio features representing interpersonal behavior in the joint model were analyzed. Again, the relative importance scores of both facial and audio features in the joint model were computed. The score comparisons indicate that this model relies more heavily on facial expressions (93.6%) to evaluate interpersonal perception between the interactants compared to audio features (6.4%). Similarly, analyses on ego (facial: 89.6%, audio: 10.4%) and partner (facial: 89.5%, audio: 10.5%) models found that facial features are dominant over audio features in case of the individual models. Based on these findings, facial features with the highest importance scores were used to identify important facial regions for the ego, partner, and joint models, shown in Figure 3.

In addition, results of the feature importance analysis on facial activation units showed that head movements of both interactants carry high predictive power. The head pitch, roll, and yaw angles are among the most important features for both ego and partner. This is not surprising given that the dataset consists of virtual interactions between interactants, where head movements can be important indicators of attentional focus towards the interaction partner. Similar analyses for both ego and partner models were performed, where, once again, head movement-related action units were found to carry high predictive power.

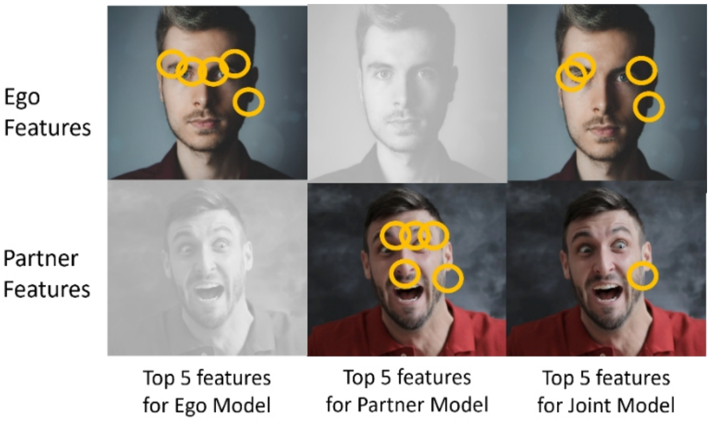

Figure 3: A comparison of important facial regions in the ego, partner, and joint models. For visualization purposes, only the top five features from each model are displayed.

Negative and Positive Labels

The experiment aimed to address RQ3 by identifying differences in the importance of audiovisual cues in positively and negatively rated interaction segments. The results of this experiment offer insights into the differences in behavioral expressions and modalities utilized by the interactants to express positive and negative perceptions of their partners. From the feature importance analysis on positively rated dyadic interaction data, it was found that the importance of audio features increases for the model predicting negative interpersonal affect (facial: 82.1%, audio: 17.9%) as compared to the one that predicts positive affect (facial: 93.6%, audio: 6.4%). This indicates that verbal behavior may be a more important indicator of negative interpersonal affect rather than positive interpersonal affect in the given dyadic interaction dataset.

Other Findings

Motivated by prior research that suggests a difference in how the two sides of a face may express emotions (Sackeim, 1978), investigations were made to validate if affective expressions in this dataset are aligned with these findings. Literature suggests that the left side of the face may be more expressive for low to intermediate level intensity of expressions, while the right side may be more expressive for most intense expressions (Mandal, 1995). To validate, feature importance analysis was conducted to evaluate the contributions of the features from each side of the interactants' faces to the joint model's performance.

From feature importance analysis on the left- and right-sided features, a higher overall importance for left-sided features was identified (left: 48.6%, nose: 12.1%, right: 39.3%). In line with these outcomes, it was found that the contribution of left-sided and right-sided features in the prediction of negative (left: 43.0%, nose: 16.0%, right: 41.0%) and positive interpersonal affect (left: 49.4%, nose: 11.8%, right: 38.8%) shows a higher importance for left-sided facial features. This understanding of the different predictive powers of the two sides of the face may inform the design of technology-assisted mediation strategies, enhancing their adaptability and effectiveness in scenarios where obstructions to the interactants' faces are challenging to mitigate.

DISCUSSION

This work investigated the behavioral expressions that reflect interpersonal affect in dyadic interactions within the given dataset. The findings from this work offer insights into how mediation systems may benefit from multi-perspective approaches for interpersonal affect evaluation.

The results from experiment 1 show that representing both interactants in the feature space can capture the interplay of behaviors between the interactants, such as eye gaze, tone of voice, nonverbal backchanneling, etc. This exchange of information through behavioral expressions is inherent to human-human interactions that may not be fully represented by individual models alone. It is also found that the behavioral evidence of interpersonal affect is more pronounced in features of the interactant rating their partner rather than the partner being rated. This understanding underscores the importance of considering reciprocal perceptions in modeling human interactions, especially in real-world scenarios where real-time processing constraints necessitate prioritizing one participant's features to enable more efficient modeling of interpersonal dynamics.

From experiment 2, it is found that head movements are important behavior cues for predicting interpersonal affect in this dataset. The importance of left-sided features was compared to right-sided features, where the left side of the face was found to produce features that are more important for modeling interpersonal affect. Experiment 3 suggested that audio features may be better indicators of negative interpersonal affect than positive. These insights offer a deeper understanding of the dyadic interactions represented in the dataset, which can be used to create more robust mediation technologies for deployment in the wild.

Real-world deployment presents challenges for mediation technologies, such as obstructions to the face, variable lighting, overlapping speech, etc. These issues dramatically increase the difficulty of estimating interpersonal affect due to partial data availability. In addition, the mediator may be required to process large amounts of multimodal behavioral data from multiple interactants, with additional time constraints on generating effective mediation strategies. Therefore, a deeper understanding of dyadic human interactions and the various factors that can influence the predictive power of behavioral data may be beneficial for the design of robust mediation technologies. The understanding of important facial regions in the interactants could enable a mediator to produce reliable estimates of interpersonal affect even when access to some facial regions may be limited. Knowledge of the contributions of audio and facial features toward model performance may be leveraged to reach reliable estimates of interpersonal perception when the quality of one modality is low. Comparison of the importance of features from the left- and right-side of the face may influence how data collection methods are designed to favor angles that yield more important features.

These findings also shed light on the underlying interaction dynamics contained in the dataset, explaining the contribution of the different modalities and interactants involved. They may also inform modeling approaches that utilize decision-fusion (Bota, et al., 2020) or modality adaptation (Razzaghi, et al., 2021), offering insights to leverage the different predictive powers of each modality.

While further investigation is required to validate the generalizability of these findings to other human-human interaction datasets, this work takes preliminary steps towards a multi-perspective approach for interpersonal affect estimation in human-human interactions. Future work could also investigate how interpersonal dynamics may vary for interactions between larger groups. Incorporating long-term audio features may also help capture relationship dynamics that current short-term descriptors cannot. Modeling approaches that leverage the temporal dependencies in the behavioral data will also be important to explore in the future.

CONCLUSION

This work investigated how behavioral cues can be best employed for technology-assisted mediation. Findings suggested that an interactant's self-behavior may be more reflective of interpersonal perception than their partner's behavior. The results underscore the need to employ behavioral data from both interactants, informing which self and partner features to leverage, to produce effective mediation strategies. An understanding of the interaction from the perspective of each individual interactant enables unique and targeted mediation methods that can be tailored to address specific interpersonal dynamics and foster more effective and harmonious social interactions. The multi-perspective computational method proposed in this work lays the groundwork for mediations that can personalized to each interactant's needs.

APPENDIX A

List of raw, 2D facial landmarks used in this work:

Right top jaw
right jaw angle
gnathion
left jaw angle
left top jaw
outer right brow corner
right brow center
inner right brow corner
Inner left brow corner
left brow center
outer left brow corner
nose root
nose tip
nose lower right boundary
nose bottom boundary
nose lower left boundary
Outer right eye
inner right eye
inner left eye
outer left eye
right lip corner
right apex upper lip
upper lip center
left apex upper lip
Left lip corner
left edge lower lip
lower lip center
right edge lower lip
bottom upper lip
top lower lip
upper corner right eye
lower corner right eye

List of short-term audio features used in this work:

ZCR
spectral roll-off
energy entropy
spectral centroid
spectral spread
spectral entropy
Spectral flux
energy
MFCC 1
MFCC 2
MFCC 3
MFCC 4
MFCC 5
MFCC 6
MFCC 7
MFCC 8
MFCC 9
MFCC 10
MFCC 11
MFCC 12
MFCC 13
chroma 1
chroma 2
chroma 3
Chroma 4
chroma 5
chroma 6
chroma 7
chroma 8
chroma 9
Chroma 10
chroma 11
chroma 12
chroma std

List of facial activation units used in this work:

Pitch
Valence
Roll
Yaw
Anger
Smirk
Lip Pucker
Eye Widen
Brow Furrow
Lid Tighten
Lip Corner Depressor
Cheek Raise
Chin Raise
Brow Raise
Engagement
Contempt
Mouth Open
Eye Closure
Jaw Drop
Inner Brow Raise
Upper Lip Raise
Nose Wrinkle
Lip Stretch
Lip Suck
Attention
Fear
Sadness
Joy
Dimpler
Surprise
Lip Press
Smile
Disgust

ACKNOWLEDGMENT

This work was completed when Nina Moorman was an intern at Honda Research Institute USA, Inc.

REFERENCES

Bota, P., Wang, C., Fred, A. & Silva, H., 2020. Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet? *Sensors.*

Busso, C. et al., 2004. *Analysis of emotion recognition using facial expressions, speech and multimodal information.* Proceedings of the 6th international conference on Multimodal interfaces.

Ekman, P. a. F. W., 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior.*

Forsyth, D. R., 2014. *Group dynamics.* Wadsworth Cengage Learning.

Jung, M. F., 2017. *Affective grounding in human-robot interaction.* Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction.

Khoshgoftaar, T. M., Golawala, M. & Van Hulse, J., 2007. *An empirical study of learning from imbalanced data using random forest.* 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007).

Mandal, M. K. a. A. H. S. a. P. R., 1995. Asymmetry in emotional face: Its role in intensity of expression. *The Journal of Psychology,* Volume 129, pp. 235–241.

Razzaghi, P., Abbasi, K., Shirazi, M. & Shabani, N., 2021. Modality adaptation in multimodal data. *Expert Systems with Applications.*

Sackeim, H. A. a. G. R. C. a. S. M. C., 1978. Emotions are expressed more intensely on the left side of the face. *Science,* Volume 202, pp. 434–436.

Sack, M. J. et al., 2023. *CORAE: A Tool for Intuitive and Continuous Retrospective Evaluation of Interactions.* 11th International Conference on Affective Computing and Intelligent Interaction (ACII).

Shek, D. T., 2002. Chinese adolescents' explanations of poverty: the Perceived Causes of Poverty Scale. *Adolescence.*

Van Kleef, G. A. & Côté, S., 2022. The social effects of emotions. *Annual review of psychology,* pp. 629–658.

Veltmeijer, E. A., Gettisen, C. & Hindrikis, K. V., 2021. Automatic emotion recognition for groups: A review. *IEEE Transactions on Affective Computing.*