Tilburg University

UNDERSTANDING THE IMPACT OF CONVERSATION SENTIMENT AND EMOTION ON SUBJECTIVE TEAM PERFORMANCE IN DYADS

FAHAD SHAKEEL

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBER

2044393

COMMITTEE

dr. Travis Wiltshire
dr. Eriko Fukuda

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

June 24, 2022

ACKNOWLEDGMENTS

I would like to thank my supervisor, dr. Travis Wiltshire, for the provision of the dataset used in this study, and for the provided guidance and expertise throughout the entirety of the thesis.

UNDERSTANDING THE IMPACT OF CONVERSATION SENTIMENT AND EMOTION ON SUBJECTIVE TEAM PERFORMANCE IN DYADS

FAHAD SHAKEEL

Abstract

Due to the increasing complexity of different tasks and challenges, the past few years have shown increasing demand for teamwork and collaboration. This study investigates teams by examining the interactions between components of the team instead of their individual actions. The EATMINT dataset from Chanel, Bétrancourt, Pun, Cereghetti, and Molinari (2013) was used to conduct this research. The dataset consisted of multimodal recordings of 30 dyads (teams of two) collaborating on creating a slogan against school violence. The first goal of this study was to investigate the impact of dyadic conversational sentiment on subjective team performance. The second goal of this study was to examine the relationship between the emotions perceived by participants and the physiological data recorded from the task. For both parts, multiple linear regression models were generated. The results for the first part showed that conversation sentiment significantly predicted two out of eight features of the dyad's collaboration. For the second part, no significant impacts were observed; nevertheless, certain types of emotions had a stronger effect on the physiological signals than others. This study added to team dynamics research by evaluating the impact of conversation sentiment on team performance in dyads and by locating patterns in emotion and physiological data in the context of teamwork. Future studies could use different measures of physiological data than those used in this study, to improve understanding of established links with perceived emotions during a collaborative task.

1 INTRODUCTION

Numerous challenges in today's world have become too complicated to be solved by individuals. The growth in problem complexity could be due to factors like time restraints or the interdisciplinary character of the problems. Consequently, there has been an increase in the need for collaboration between individuals in completing tasks and solving challenges. To effectively address such issues, it is critical to understand the factors influencing individuals in a team setting to collaborate more or less effectively. Therefore, a deeper understanding of the fundamental principles governing team coordination dynamics would be helpful. These principles could allow us to identify and control the significant predictors of performance, potentially benefiting every area of development involving collaboration. Since teamwork is involved in many tasks and activities, understanding team dynamics and factors affecting performance could advance fundamental areas of development, like healthcare, businesses, and education.

A more refined comprehension of the impact of individual interactions on the output quality of a system would equip us with a toolbox that can be utilized to moderate different aspects of a team setting, such as productivity and chemistry. If team performance prediction is capable of differentiating between various components of a team, more appropriate intervention techniques may be used to fasten the process of improving team performance and the quality of output. The ability to isolate the problematic component could further allow for the implementation of target-specific intervention measures that maximize expected performance. Therefore, this study examines teams from the perspective of dynamical systems theory, as stated by Gorman, Dunbar, Grimm, and Gipson (2017): instead of concentrating exclusively on the conduct of individual team members, team dynamics are investigated through their interactions. The dynamical systems theory states that synchronization occurs when two or more different components of a dynamical system interact. This concept could also be extended to teams, where individuals influence one another as a result of their informational coupling.

The goal of this research is to understand factors affecting subjective team performance during a collaborative task. The data required for this study will be taken from the EATMINT dataset (Chanel et al., 2013). The first part of this study uses the sentiment of the conversation between team members collaborating on a task to predict subjective team performance. This would allow us to understand how the conversation sentiment impacts each component of the team output. The second part of this study examines how interpersonal physiological data is related to the perceived emotions reported by the participants after completing a collaborative task in teams of two. The second part of this study also aims to understand what kind of emotions are related to specific physiological measures and uncover patterns, if any. The physiological signals used in this study are electrocardiogram (ECG) and electrodermal activity (EDA). Previous research studies examining interpersonal collaboration in teams, like Ahonen et al. (2016) and Stuldreher, Thammasan, Van Erp, and Brouwer (2020), have also used the physiological signals EDA and ECG, sometimes with an additional multimodal metric. Generally, this study aims to contribute to the field of team dynamics by answering questions that have not been addressed in this context previously.

Previous studies have conducted research with single-modality (Kazi et al., 2021) and physiological data (Ahonen, Cowley, Hellas, & Puolamäki, 2018; R. Henning, Armstead, & Ferris, 2009; R. A. Henning, Boucsein, & Gil, 2001) to predict team performance. However, no such literature was found where conversation sentiment was considered a driver of change within the notion of team dynamics and dyads. Communication has shown to be the most important factor influencing performance in collaborative problem-solving situations (Yost & Tucker, 2000), and sentiment is a crucial aspect of communication. To this end, the following research question was generated:

> *RQ1: Does positive conversation sentiment during a collaborative task predict subjective team performance in dyads?*

Positive sentiment in teams influences decision-making and information tactics used by the team members (Chartrand, van Baaren, & Bargh, 2006). Also, individuals process negative or positive sentiment words more quickly than neutral words (Kousta, Vinson, & Vigliocco, 2009). Results from the EATMINT (Emotion Awareness Tools for Mediated Interaction) project indicate that perceived transactivity and group performance exhibit a positive correlation with positive emotions such as enjoyment and negative correlation with negative emotions such as boredom (Molinari & Avry, 2018). Given these findings, the following hypothesis was drawn:

> *H1: Higher (more positive) sentiment score leads to better team performance in a collaborative task.*

Numerous studies in the past have already found a correlation between team performance and physiological coordination (Ahonen et al., 2016; R. A. Henning et al., 2001; Walker, Muth, Switzer III, & Rosopa, 2013). However, the relation between interpersonal physiological data and perceived emotion from a collaboration has not been examined before. Investigating the relationship between physiological coordination and perceived emotion could offer interesting insights regarding the possibility to enhance the current state-of-the-art emotion recognition models for collaboration settings. It could also be useful to understand how much the intensity of a particular type of emotion can explain the variation in physiological signals. Therefore, the following is the second research question of this study:

> *RQ2: How does interpersonal physiological data relate to perceived emotions reported by participants after a collaborative task in dyads?*

2 RELATED WORK

2.1 *Previous studies using the EATMINT dataset*

The EATMINT dataset was created by Chanel et al. (2013) in a study where they used physiological signals and eye movements data to evaluate essential collaboration factors during remotely supported group work. They trained two regression models to evaluate cooperation based on extracted features from the collected time series that measure the extent to which collaborators' eye movements and physiology are related. The findings suggested that the coupling metrics could be used for forecasting collaborative processes such as grounding and convergence. They eventually evaluated that assessing these processes is a crucial step in developing remote collaboration interfaces capable of adapting to the users' social interactions.

Another important by Chanel, Avry, Molinari, Bétrancourt, and Pun (2017) worked with the same dataset to study emotion recognition in the context of teamwork. The aim was to investigate whether the feelings of the participants could be deduced from their partners' emotional reactions and behaviors. The study of two forms of emotional displays, physiological responses and speech, was conducted. The results indicated that by using affective characteristics from either the self or the other, emotions may be identified with comparable accuracy. In addition, improved performance was observed when self-information and relationship information were combined. The results indicate that an emotion recognition model should contain information about partner in social settings.

Since the first study focused on subjective team performance from the perspective of physiological and eye-movement data and the second study was based on emotion recognition, there is still a need to research and analyze other potential factors, such as conversation sentiment. This thesis would be able to evaluate if the emotion recognition model suggested by Chanel et al. (2017) would benefit from the addition of heart rate variability and electrodermal activity measures.

2.2 *Related work on text sentiment*

Communication between individuals can be analyzed and interpreted by observing the sentiment of the conversations. An individual's feelings toward an entity can be defined as sentiment. A person's feelings, cognition, and attitudes largely influence sentiment (B. Liu, 2012). A sentiment analysis is conducted to research and understand the sentiment of some text. Sentiment analysis (or opinion mining) is a natural language processing (NLP) technique that allows for the study of people's opinions, sentiments, emotions, and attitudes towards various entities (P. Liu, Joty, & Meng, 2015). Sentiment analysis uses polarity measures to classify data (B. Liu, 2012), for example, with three polarity categories, text data can be categorized as neutral, positive, or negative. Previous studies have employed this analysis technique on tweets and product reviews (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011; Kotzé & Senekal, 2018; Saif, He, & Alani, 2012; Wang, Can, Kazemzadeh, Bar, & Narayanan, 2012) to gauge the opinions of people.

There are two main approaches for sentiment analysis as categorized by Medhat, Hassan, and Korashy (2014): machine learning-based and lexicon-based. The former usually utilizes supervised machine learning algorithms like Naive Bayes and Support Vector Machines to extract features from the data and detect patterns. It allows to make decisions under uncertainty or forecast future data (Abirami & Gayathri, 2017). On the other hand, the latter approach is based on machine translation that uses particular dictionaries, where each word is assigned a score indicating its polarity category (negative, neutral, or positive) (B. Liu et al., 2010). However, a machine learning approach requires training the model using data that consists of texts annotated with their polarities.

Due to the efficiency of Part of Speech tagging (POS) algorithms in lexicon-based approaches, sentiment analysis techniques based on lexicon are better performing than machine learning-based tools (Rhouati, Berrich, Belkasmi, & Bouchentouf, 2018). Also, for training a machine learning model, a corpus most representative of the task needs to be found. Therefore, this study followed a lexicon-based analysis approach using french support offered by the TextBlob library (Loria, 2018) in Python (Van Rossum & Drake, 2009). This library is built on the Natural Language Toolkit (NLTK) (Bird, Klein, & Loper, 2009), which it uses to achieve sentiment analysis tasks. TextBlob has been used in recent studies like Gujjar and Kumar (2021) to conduct sentiment analysis tasks.

2.3 *Related work on teamwork and collaboration*

Teamwork can be defined as the collaborative effort of two or more individuals working toward a common objective. Earlier attempts to comprehend effectiveness in teams tended to concentrate on how team factors (such as individual and team qualities) influenced team outcomes (e.g., performance, content) (Goldstein & Ford, 1993; Guzzo & Dickson, 1996). However, these models did not offer a detailed understanding because they could not account for the dynamic, moment-to-moment actions and complex interactions among team members during task completion. Collaboration is a complicated phenomenon in which intersubjective dynamics can have a significant effect on the output. Thus, the assessment of collaboration is of immense value and has the potential to contribute to improved outcomes and performance (Ahonen et al., 2018). However, quantitative evaluation of collaboration is challenging due to the intersubjective nature of the interactions between team members. Chanel et al. (2013) worked with collaboration between dyads and suggested that the quality of collaboration between team members can be assessed by either analyzing the collaborative processes involved or the team performance. In contrast to performance evaluation, collaborative processes evaluation identifies when and which adaptive feedback is required. These collaborative processes can be evaluated through a self-report questionnaire asking participants about their interaction with their fellow team member.

Numerous processes that significantly influence successful collaboration have been found. Examples of such processes include maintaining shared understanding (grounding), constructing a representation of the team members (mutual modeling), co-construction of knowledge, developing upon the team member's reasoning (transactivity), specifying and managing conflict, mutual decision making and convergence, and regulation of interdependence. Such interaction dynamics are known to be positively associated with individual results, such as learning gains, and team performance (Meier, Spada, & Rummel, 2007; Weinberger, Stegmann, & Fischer, 2007).

2.4 *Related work on emotion*

In social interactions, emotions do not form only in one's mind, but rather emerge in response to the emotional responses of others in close proximity (Boiger & Mesquita, 2012). Similarly, it is well established that behaviors, expressions, and physiology of interacting persons are synchronized throughout interactions (Bilakhia, Petridis, Nijholt, & Pantic, 2015; Chanel & Mühl, 2015). Van Kleef, Homan, and Cheshin (2012) revealed that the opportunity of identifying emotional expressions that impact another person's behavior develops during teamwork. This viewpoint is compatible with the notion that emotions may be directed toward other individuals. It was found that team coordination is influenced by the emotions among members. Furthermore, intricate circumstances require teams to depend more on affective components, such as team emotions (Forgas & George, 2001).

These findings motivate studying emotions (more than other factors) in team settings where interaction and collaboration are of significance. Moreover, evaluations regarding emotion are expected to be different in individual situations than during collaboration with another individual. Given the subjectivity of emotion, it is often difficult to evaluate them, and for this reason, scholars appear to have differences in their interpretation of emotion. To better describe human emotions, researchers have established dimensional spaces, which highlight the similarities and distinctions between different emotional experiences. However, the two-dimensional circumplex model from Russell (1980) is the most extensively used dimensional model in emotion research. This model uses valence and arousal as the axes of the model. Therefore, Russell's circumplex model is used in this study to sample emotions for analysis (Section 3.5).

2.5 *Related work on interpersonal physiological data*

This study will investigate two types of physiological signals in relation to emotion: 1) electrodermal activity (EDA) and 2) heart rate variability (HRV), which was obtained from electrocardiogram data. HRV refers to the variation in the time between heartbeats (Mather & Thayer, 2018) and EDA refers to the autonomic variations of the skin's electrical properties (Braithwaite, Watson, Jones, & Rowe, 2013). The motivation for selecting HRV and EDA specifically, is based on findings from previous studies.

A higher HRV has been linked with higher psychological and emotional well-being (Kemp & Quintana, 2013; Shaffer, McCraty, & Zerr, 2014), while also being associated with typically more controlled emotional response (Appelhans & Luecken, 2006). Therefore, individuals with a high HRV tend to be better at emotional regulation. A research study by Critchley (2002) found EDA to be strongly associated with emotional arousal. Higher EDA activity has also been linked to a range of emotional states, indicating the significance of the physiological response in the emotional experience (Kreibig, 2010). However, studies like this one were conducted with individual participants and thus, the link between EDA and emotions perceived in a team collaboration setting still needs to be evaluated.

Research studies like Ahonen et al. (2016) and Walker et al. (2013) have found relations between physiological coordination and team performance. These findings are also well-established in the context of this study since most of the studies were conducted with dyads. Even studies examining performance in collaboration settings with groups of more than two split the groups into groups of two for the analysis part (Elkins et al., 2009; R. Henning et al., 2009). Since this study examines the relationship between perceived emotions and different physiological signals, the results could be able to add to the understanding of EDA and HRV from a teams perspective. Due to the variations in the interpretation of emotion, physiological signals being objective measures may act as indicators of certain emotions during collaboration.

3 METHOD

In this section, a brief description of the dataset along with the experimental setup is provided at first. Next, the methodology employed for the preprocessing and analysis of the data is discussed in detail. Finally, a brief overview is given of the statistical tests conducted to answer the research questions and test the hypothesis. Most of the data cleaning and manipulation was completed using the Pandas library (McKinney et al., 2010) in Python (Van Rossum & Drake, 2009).

3.1 *Dataset*

This study makes use of data taken from the publicly accessible EATMINT dataset, as reported by (Chanel et al., 2013). The EATMINT dataset is comprised of multimodal recordings of 30 dyads (teams of two people), where dyad members were of the same gender. A total of 16 women dyads and 14 men dyads (mean age = 23.5) participated in a task where they were instructed to create slogans against school violence. The collaborative task had to be completed by the dyad in less than 45 minutes. These dyads collaborated remotely via headphones using the collaborative environment DREW, without being able to see each other. The first portion of the experiment required the collaborating dyads to come up with as many slogan concepts as possible. The second portion of the experiment required dyads to argue the appropriateness of their slogan suggestions and eliminate the less relevant ones. The last portion of the experiment required dyads to reach a conclusion on the ideal slogan. After the experiment, the participants were asked to answer a questionnaire regarding their perception of the task and the interactions within. The questionnaire was divided into four groups of questions and evaluated on 7-point Likert scales. These questions pertained to participants' perceptions of their connection with their team members, the amount of time spent on collaborative tasks, the frequency with which they exhibited a particular behavior, and the frequency with which their partner exhibited a particular behavior.

Participants' actions, facial gestures, speech, eye movements, and physiological signs were all captured during their participation. The participants were allowed to exhibit any social or emotional behavior they pleased. Recordings from both participants in the dyads are synchronized, which enables an easy combination of data from both individuals. Furthermore, the dataset contains eight variables that characterize various aspects of each participant's perceived collaboration: Grounding & Coordination, Emotion Modeling, Confrontation & Consensus Building, Co-construction, Degree of Convergence, Degree of Conflict, Emotion Management, and Transactivity. Throughout this paper, these collaboration factors will be as mentioned as 'Grounding', 'Emotion Modeling', 'Confrontation', 'Co-construction', 'Convergence', 'Conflict', 'Emotion Management', and 'Transactivity', respectively. Since the dataset also contains emotion annotations, physiological data (electrodermal activity and electrocardiogram signals), and conversation transcripts between individual participants of a dyad, it was a great fit for this study.

3.2 *Sentiment Analysis*

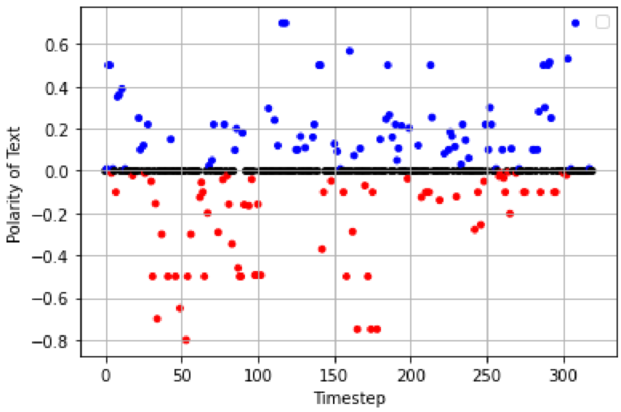
The first research question of this study required the retrieval of sentiment from the participants' conversations. To this end, sentiment analysis was conducted. The conversation transcripts for the dyads, found in the dataset, were written in the French language and were only made for 19 out of the 30 dyads. To avoid the loss of information and uncertainty in analysis, the texts were not translated into English. However, compared to English, there are fewer resources available for the French language and many libraries and tools do not offer support for analyzing French text. For this reason, the options for approach selection were limited.

The key steps involved in sentiment analysis using TextBlob are tokenization, lemmatization, and POS tagging. Tokenization refers to the breakdown of text into a list of individual components called tokens (could be a word, punctuation, or a special character). Lemmatization refers to the process of grouping together words that have the same root or lemma but varying inflections or meaning derivatives so that they may be analyzed as a single unit. For example, lemmatizing the words "dogs", "dog's", and "dogs'" involves removing the suffixes "s", "'s", and "s'" to reveal the underlying word "dog". POS tagging is the process of assigning a grammatical category to the tokens in the text. Examples of tags include ADJ (adjective), CCONJ (coordinating conjunction), and NOUN.

Each text in the conversation between members of a dyad was assigned a polarity score and category. A text could only be categorized as positive, negative, or neutral, and the polarity of a text could range between -1 and +1. A neutral text was one with a polarity score of 0; any text with a higher score than 0 was classified as a positive text, and a text with a lower score than 0 was classified as a negative text. This particular threshold was selected due to its use in recent studies like Gujjar and Kumar (2021) that worked with TextBlob. Dyads were not assigned a sentiment category since the overall sentiment for approximately all conversations was positive. Instead, a mean polarity score was assigned to each dyad, which was calculated by averaging polarity scores for all texts in the conversation.

For each dyad, there were the three measures derived: mean polarity score, number of positive texts in the conversation, and the number of negative texts in the conversation. Figure 1 shows a visualization of all the texts in a randomly selected conversation, based on their polarity scores. The reason for taking the number of positive texts and number of negative texts as measures of sentiment in the analysis could be attributed to the nature of conversations. As can be seen in Figure 1, there were texts with high negative scores as there were texts with high positive scores. While taking the mean polarity from all texts provides information about the overall sentiment of the conversation, it fails to reflect the presence of highly negative or highly positive texts.

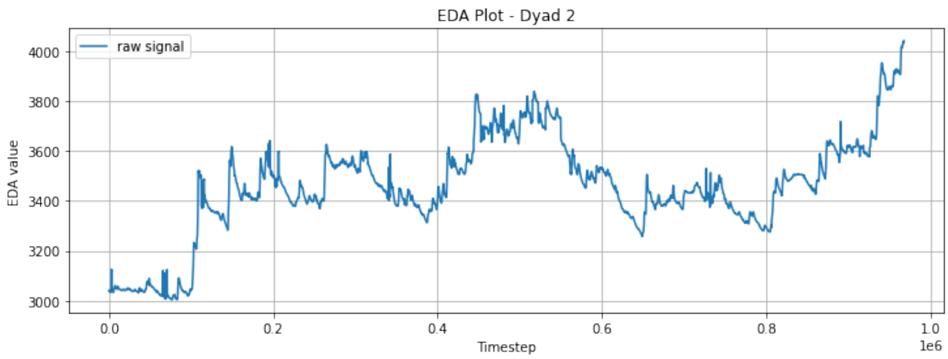
Figure 1: Polarity distribution for the conversation of a randomly selected dyad, where blue, black, and red points represent positive, neutral, and negative texts, respectively.



3.3 *Electrodermal Activity (EDA)*

The electrodermal activity (also known as galvanic skin response), sampled at 512 Hz, was analyzed using the *Neurokit2* (Makowski et al., 2021) package in Python (Van Rossum & Drake, 2009). Figure 2 shows an example of raw EDA signal from one of the participants. *Neurokit2* was used to detect Skin Conductance Response (SCR) Peaks from the raw signals. When EDA activity varies significantly in response to a stimulus, this is referred to as an Event-Related Skin Conductance Response (ER-SCR). These signals, also known as EDA peaks, can provide information regarding stimuli-induced emotional arousal (Caruelle, Gustafsson, Shams, & Lervik-Olsen, 2019). A summary analysis of emotion studies employing EDA measurements by Caruelle et al. (2019) revealed frequent use of the number of EDA peaks as a measure of EDA. However, this study uses the number of peaks per minute, and not the total number of peaks, as a measure of EDA due to the differences in collaboration time between dyads. Visualization for the peak detection of the EDA signals can be found in Appendix B (Figure 15).

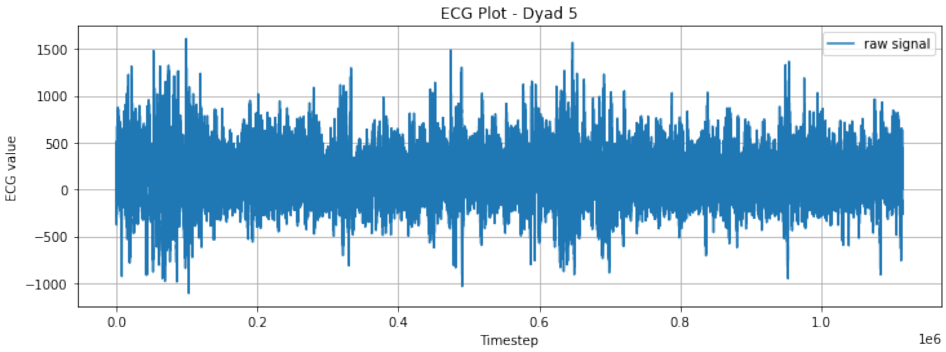
Figure 2: Raw EDA signal for a randomly selected participant from Dyad 2.



3.4 *Heart Rate Variability (HRV)*

The heart rate variability was calculated using the electrocardiogram (ECG) data, which was sampled at 512 Hz. The ECG signals were first filtered using a band-pass filter between 0.05 Hz and 40 Hz as recommended by Chanel et al. (2013). Figure 3 shows an example of the ECG signal after the band-pass filter was applied. The filtering was implemented in Python using the *HeartPy* package (Van Gent, Farah, Van Nes, & Van Arem, 2019). The R peaks were then extracted from the filtered signals using the *Neurokit2* package (Makowski et al., 2021) in Python. Figure 4 shows a zoomed-in view of the filtered ECG signal, showing the R peaks. The method followed for the R peaks detection was the Pan & Tompkins algorithm (Pan & Tompkins, 1985) integrated within *Neurokit2*.

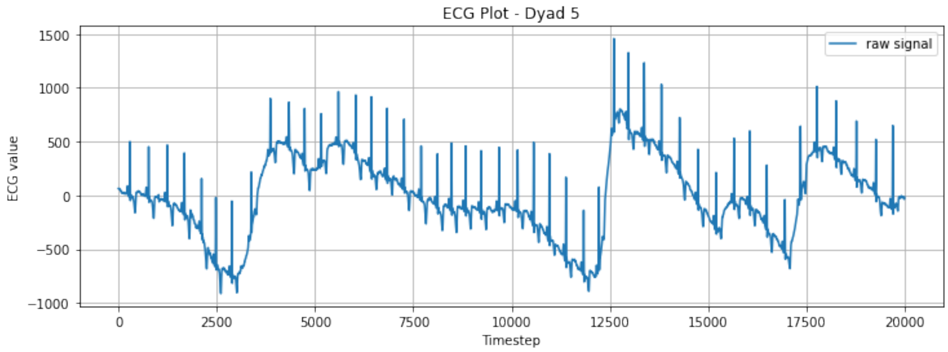
Figure 3: ECG signal after the band-pass filter between 0.05 Hz and 40 Hz was applied, for a randomly selected participant. (see Appendix B, Figure 14 for an example of the unfiltered ECG signal)



There were two measures of heart rate variability that were used in this study: Beats Per Minute (BPM) and Root Mean Square of Successive Differences between each heartbeat (RMSSD). RMSSD is the primary time-domain metric used to assess the vagally-mediated alterations exhibited in HRV (Shaffer et al., 2014). Previous studies like Mather and Thayer (2018) have used RMSSD as a measure of HRV when examining the relationship between HRV and parts of the human brain involved in the regulation of emotion.

Both of these measures were obtained from the R peaks time-series. BPM was calculated by dividing the number of R peaks by the total time of the collaboration. On the other hand, RMSSD was determined by computing the time interval between successive heartbeats in milliseconds (also known as RR interval). Then, each value is squared and the result is averaged before calculating the square root of the total (see Appendix B, Figure 12). The distribution of the RR intervals, or R peaks, can be found in Appendix B (Figure 13). Due to invalid physiological data (either from ECG or EDA), from either member of a dyad, the dyads 4 and 27 were not included in the analysis (see Appendix B: Figure 16, Figure 17, and Figure 18). Moreover, due to trigger issues, physiological signals for dyad 14 were not recorded. Therefore, only 27 dyads were included in the final analyses.

Figure 4: Zoomed-in view of the filtered ECG signal for a randomly selected participant (Dyad 5, Participant 1), showing the R peaks.

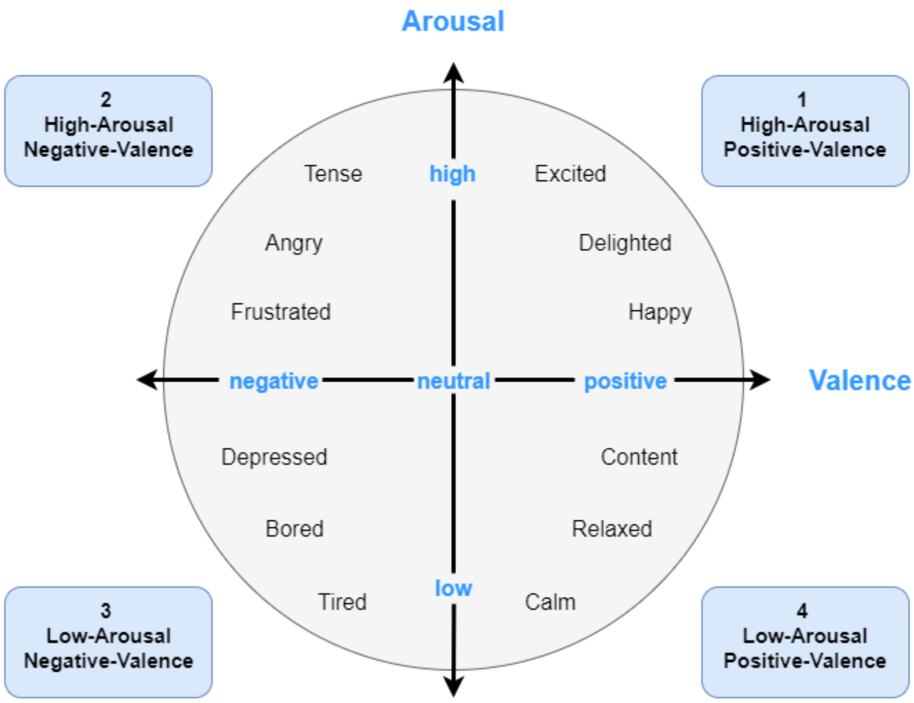


3.5 *Emotion*

Emotion data was retrieved from the questionnaire answered by the participants after completing the task. Each participant was required to describe 19 different emotions (labeled in French) on a 7-point Likert scale. However, only the following 4 emotions out of the total 19 were included in the analysis: 'Frustre', 'Ravi', 'Lasse', and 'Detendu'. Using the DeepL translator, these emotions translate to 'Frustrated', 'Delighted', 'Tired', and 'Relaxed', respectively. The complete list of emotions included in the questionnaire, along with their English translations can be found in Appendix A (Figure 7). After sub-setting the data for the four emotions, there were 2 null values found for the 'Lasse' emotion. The null values were replaced with a 0 since any higher value would indicate that the emotion had been perceived by the participant with some intensity. As 0 does not lie on the 7-point Likert scale, it represents that a certain emotion was not perceived by the participant.

The sampling of emotions was based on the circumplex model (Russell, 1980). All emotional experiences, according to the circumplex model, are spread in a circular space supported by the two perpendicular axes, valence and arousal. Arousal (or intensity) is the level of autonomic activity elicited by an event, ranging from peaceful (or low) to aroused (or high), and valence is the event's level of pleasure and satisfaction and is described along a continuum ranging from negative to positive (Bestelmeyer, Kotz, & Belin, 2017). Figure 5 depicts an example of a group of basic emotions established using the two-dimensional circumplex model. Each of the 4 selected emotions belonged to a different quadrant in the circumplex model. The motivation for sampling only 4 out of the 19 emotions was to represent all categories of emotion in the analysis without having a large number of predictors. The emotions Frustre, Ravi, Lasse, and Detendu were taken from quadrants 2 (High-Arousal, Negative-Valence), 1 (High-Arousal, Positive-Valence), 3 (Low-Arousal, Negative-Valence), and 4 (Low-Arousal, Positive-Valence), respectively. Due to the excluded physiological data, emotion information for the dyads 4, 14, and 27 were discarded as well.

Figure 5: The circumplex model showing the distribution of core emotions (adapted from Z. Liu et al. (2018)).



3.6 *Subjective Team Performance*

The eight collaboration factors reported by (Chanel et al., 2013) were used as the measures of subjective team performance. These eight standardized factors were obtained from the factorial analysis conducted on the four groups of questions stated in Section 3.1. All participants in the study had individual scores for each collaboration factor. To obtain a single score for each dyad, across all factors, the scores for the members of the dyads were averaged. Each collaboration factor along with its description can be found in Table 1. For each factor, the items are the variables from which the factor extracts sufficient variance, and the loadings are the correlation coefficients for the respective item and the factor. The factor loadings indicate the proportion of variance explained by an item on a certain factor.

Table 1: The collaboration factors used as measure of subjective team performance (used with the permission of Chanel et al. (2013)).

Factor Name	Description of main items related to each factor (loadings in parenthesis)
Grounding & Coordination	Maintaining a shared understanding (.88); managing the progress of the task (.80); managing the quality of the relation (.75); providing/asking for clarification (.73)
Degree of Conflict	Relational conflict (.83); conflict of ideas (.79); competition (.62); emotional tension (.60)
Degree of Convergence	Action synchrony (.77); mutual understanding (.74); conceptual convergence (.72); emotional convergence (.61); symmetry in roles and responsibilities (.68)
Confrontation & Consensus building	Discussing about disagreements (.82); defending and arguing ideas (.80); confronting different points of view (.73); negotiating and finding compromises (.68)
Co-Construction	Building together new ideas (.88); deepening and broadening ideas (.69); co-elaborating of ideas (0.67)
Emotion Management	Communicating on the emotions of others (.88 & .79); communicating on one's own emotions (.68 & .75); adapting to the emotions of others (.50 & .66); partner's effort to understand his/her own emotions (.72); partner's effort to understand emotions in others (.61)
Emotion modeling	Comparing emotions (.90 & .77); imagining reactions to emotions (.83 & .61); participant's effort to understand emotions in others (.61); participants effort to appear able to control his/her own emotions (.66)
Transactivity	Defending and arguing ideas (.74 & .61); understanding the partner's point of view (.57 & .77); providing points of view (.65 & .53); referring and building upon the partner's ideas (.60 & .55)

3.7 *Statistical Tests*

All the statistical tests were conducted in R (R Core Team, 2016) using RStudio (RStudio Team, 2015). However, some of the visualizations for the results were made using the Python library *Matplotlib* (Hunter, 2007).

To answer the first research question, multiple linear regression models were generated. Eight collaboration factors were functioning as measures of subjective team performance, and therefore, eight multiple linear regression models were generated (one for each collaboration factor as the dependant variable). The predictors used in each of these regression models were: 1) mean polarities for the conversations, 2) the number of positive texts in the conversations, and 3) the number of negative texts in the conversations. The three predictors were checked for multicollinearity and a highest correlation of .62 was found between the number of positive and negative texts. Therefore, no predictors were removed. The collaboration factors data was available for all 30 dyads, but since the transcripts were recorded for only 19 dyads, data for 11 dyads had to be excluded from the regression analysis. This could be a reason for obtaining relatively less accurate values in the results.

To answer the second research question, a total of three linear regression models were built. Each regression model used the selected emotions (Ravi, Frustre, Lasse, and Detendu) as predictors and a measure of either electrodermal activity or heart rate variability as the outcome variable. Two of the three models were built for investigating the relationship between heart rate variability and selected emotions, using RMMSD in the first and beats per minute in the second. The third model was built for observing the relationship between electrodermal activity and selected emotions, using EDA peaks per minute as the dependent variable.

4 RESULTS

To answer the first research question **"Does positive conversation sentiment during a collaborative task predict subjective team performance in dyads?"**, results from the multiple linear regression models explained in Section 3.7 were interpreted. Descriptive statistics of the collaboration factors, which function as measures of subjective team performance, were also generated (Table 2)

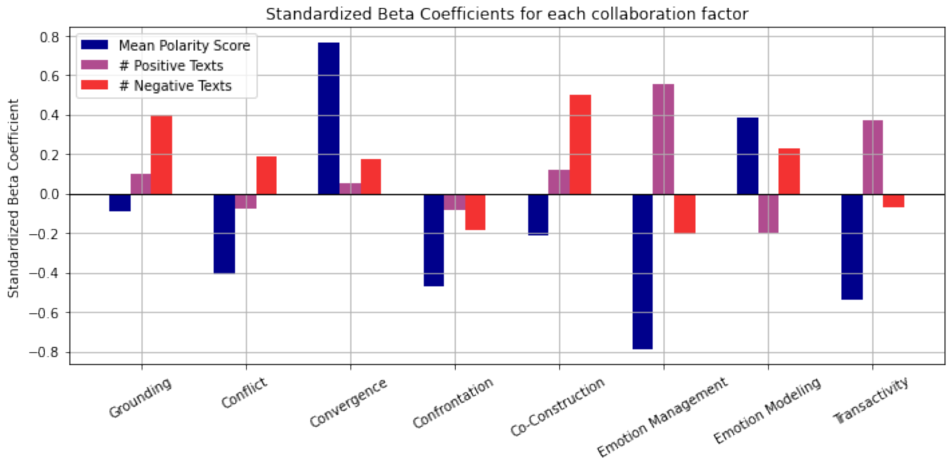
Table 2: Descriptives of the collaboration factors, after removing dyads for which transcripts were not made.

Collaboration Factor	**Mean**	**Standard Deviation**	**Range**
Grounding & Coordination	0.16	0.63	-1.46 - 1.39
Degree of Convergence	0.03	0.56	-0.93 - 1.16
Emotion Modeling	0.16	0.74	-1.28 - 2.02
Transactivity	0.03	0.61	-0.71 - 1.26
Emotion Management	0.04	0.70	-1.10 - 1.44
Degree of Conflict	-0.02	0.74	-0.79 - 2.18
Confrontation & Consensus Building	0.03	0.43	-0.54 - 1.19
Co-Construction	0.05	0.76	-1.28 - 1.15

Two out of the eight collaboration factors were predicted significantly from the regression models. For the regression model with Convergence as the outcome variable, mean polarity score was able to significantly predict 50.4% of the variability ($Adj.R^2 = 0.504, p < 0.01$). Mean polarity score was also able to significantly predict 31.7% of the variability in Emotion Management ($Adj.R^2 = 0.317, p < 0.05$). Although the other regression models did not show a significant prediction of the collaboration factors, it is possible to compare how strong the individual effects were while predicting the collaboration factors.

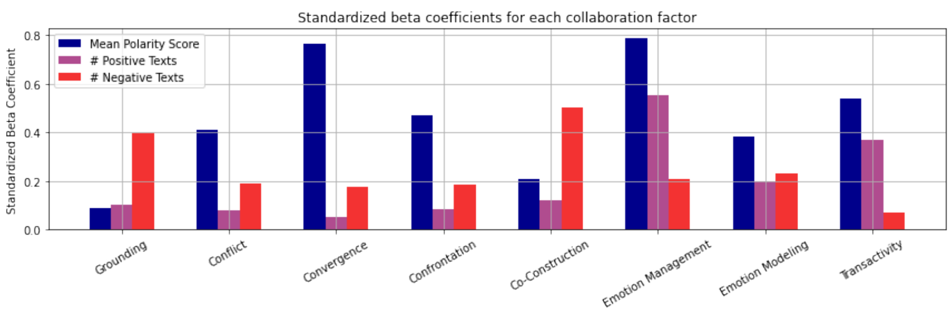
To this end, standardized beta coefficients were generated and visualized using a bar chart (Figure 6). Given the results, the hypothesis **"Higher (more positive) sentiment score leads to better team performance in a collaborative task"** could not be rejected as a higher sentiment was able to significantly predict Convergence with a high positive effect ($\beta = 0.77$). However, an interesting revelation of the analysis was the significant prediction of Emotion Management from sentiment, with a strong negative effect ($\beta = -0.58$).

Figure 6: Bar chart comparing standardized beta coefficients for the predictor variables for each regression model, grouped by collaboration factor.



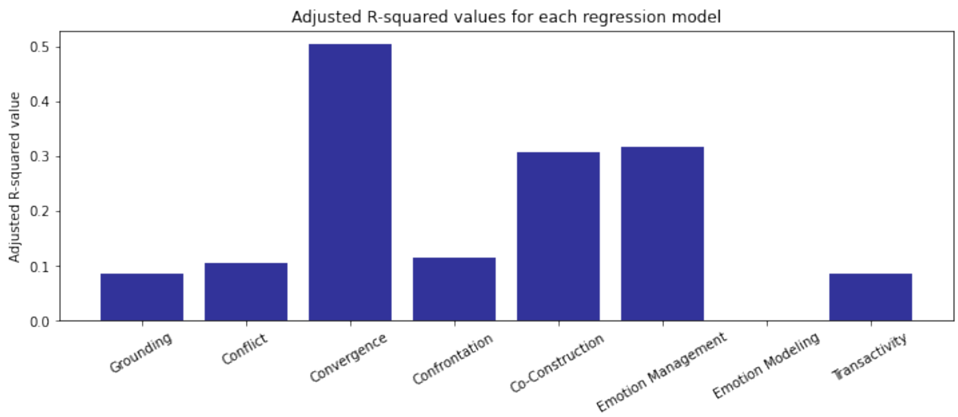
To only observe the strength of the effects (isolated from the direction), the magnitudes of the standardized beta coefficients were visualized using a bar chart (Figure 7). The results show that the mean polarity scores had the strongest effect on Convergence and Emotion Management. Even though mean polarity score predictions for other factors were not significant, the effects were relatively stronger for Transactivity, Emotion Modeling, Confrontation, and Conflict as compared to Grounding and Co-Construction. While the number of positive texts and the number of negative texts did not significantly predict any collaboration factor, these predictor variables did show relatively stronger effects for some factors. The number of positive texts had a relatively stronger effect on Emotion Management and Transactivity than all the other factors. On the other hand, the number of negative texts had a relatively stronger impact on Grounding and Co-Construction.

Figure 7: Bar chart comparing magnitude of standardized beta coefficients for the predictor variables for each regression model, grouped by collaboration factor.



The adjusted R-squared values were also compared for all the regression models to understand the differences in the variance explained by the predictors (Figure 8). The adjusted R-squared value for Emotion Modeling was negative so it was replaced with zero. The visualization shows the adjusted R-squared values for the significant regression models as the highest, however, the result for Co-Construction is close to that of Emotion Management. The results for all other models were relatively low. One possible reason for lower adjusted R-squared values is the use of multiple predictors. As discussed above, the number of positive texts and the number of negative texts only showed a somewhat strong effect on two out of the eight collaboration factors. Therefore, for the other factors, the addition of these predictors only resulted in a lower adjusted R-squared. The precise numerical values for the standardized beta coefficients and adjusted R-squared results can be found in Appendix A, in Table 3 and Table 4, respectively.

Figure 8: Bar chart comparing adjusted R-squared values for the regression models of each collaboration factor.



To answer the second research question **"How does interpersonal physiological data relate to perceived emotions reported by the participants after a collaborative task in dyads?"**, results from the three multiple linear regression models described in Section 3.7 were interpreted first. No significant predictors were observed for any of the regression models, however, the effect sizes of the predictor variables were visualized using dot-and-whisker plots. The visualizations were generated to observe the relation between the physiological data and individual emotions. A dot-and-whisker plot was generated for each regression model: root mean square of successive differences between each heartbeat (RMSSD) as the outcome variable (Figure 9), beats per minute as the outcome variable (Figure 10), and EDA peaks per minute as the outcome variable (Figure 11).

A fixed scale was used for all the dot-and-whisker plots to allow for better comparisons.

Figure 9: Dot-and-Whisker plot comparing standardized beta coefficients of the different emotions as predictors of heart rate variability, using RMSSD as the measure.

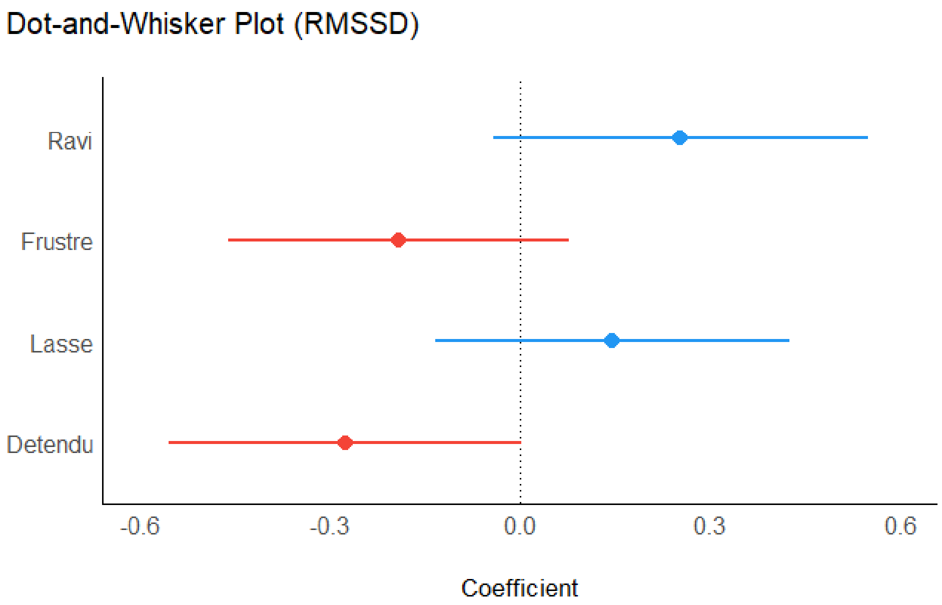


Figure 10: Dot-and-Whisker plot comparing standardized beta coefficients of the different emotions as predictors of heart rate variability, using beats per minute as the measure.

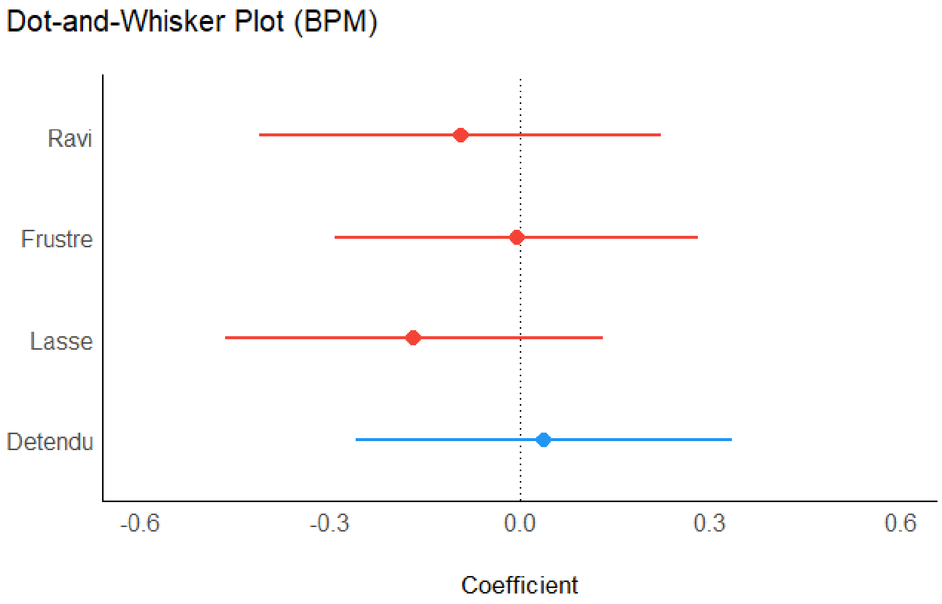
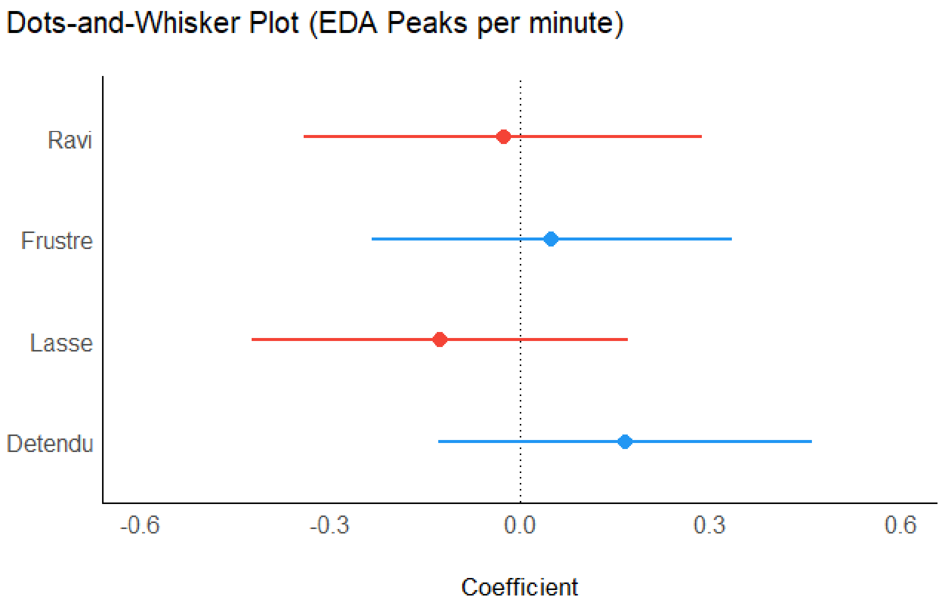


Figure 11: Dot-and-Whisker plot comparing standardized beta coefficients of the different emotions as predictors of electrodermal activity, using peaks per minute as the measure.



Since no emotion significantly predicted the three measures of physiological data, the direction of the effects and an internal comparison between them can provide important information. RMSSD was related negatively to the emotions Frustre and Detendu while being positively related to Ravi and Lasse. The strongest predictor for RMSSD, regardless of direction, was Ravi. On the other hand, Detendu had a positive effect on BPM whereas all the other emotions had a negative effect. For EDA peaks per minute as the outcome variable, Ravi and Lasse were the negatively related predictor while Frustre and Detendu were the positively related predictors. The strongest predictor for BPM and EDA peaks per minute, was Lasse. The precise numerical values for the unstandardized and standardized beta coefficients can be found in Appendix A, in Table 5 and Table 6, respectively.

5 DISCUSSION

The first goal of this study was to understand the factors influencing subjective team performance in dyads during a collaborative task. Given the previous studies conducted in this domain, it was found that the sentiment of the dyads' conversations during the collaborative task had not been considered a factor. Therefore, conversation sentiment was taken as a possible driver of change. At first, sentiment analysis was conducted to retrieve polarity scores from the text, from which three measures were derived: 1) mean polarity score of the conversation, 2) the total number of positive texts in the conversation, and 3) the total number of negative texts in the conversation. The last two measures were considered to reflect the presence of highly negative and positive texts in the conversation. Since eight different collaboration factors were taken as measures of subjective team performance, a multiple linear regression model was generated for each collaboration factor, taking the derived measures of sentiment as predictors.

The results from the regression models showed that the mean polarity score was able to significantly predict two out of the eight collaboration factors: Convergence and Emotion Management. From the significant prediction of Convergence and data from Table 2, it can be evaluated that a higher sentiment is collectively associated with improved action synchrony, mutual understanding, conceptual and emotional convergence, and the symmetry in roles and responsibilities. Interestingly, on the other hand, mean polarity score negatively predicted Emotion Management, meaning that a higher sentiment led to reduced emotion management, hence being negatively associated with the features of this factor. This collaboration factor was based on the following features: communicating on emotions of self and others, the partner's efforts to understand emotions of self and others, and general adaptability.

While there were no significant results for the other collaboration factors, the standardized beta coefficients for each model's predictors were plotted. Though the number of positive texts and the number of negative texts did not show a significant effect in any of the regression models, they were found to have relatively strong effects on some collaboration factors. The number of positive texts showed a positive effect with Emotion Management ($\beta = 0.55$) and Transactivity ($\beta = 0.37$). Additionally, the number of negative texts also showed a positive effect with two collaboration factors more than the others: Grounding ($\beta = 0.40$) and Co-Construction ($\beta = 0.50$). The adjusted R-squared values from all the models were also observed to make interpretations of the collective effect.

Conflict, Confrontation, and Emotion Modeling collectively showed the smallest associations with the predictor variables. This suggests that collaborative processes like conflict of ideas, competition, emotional tension, defending and arguing ideas, confronting different point of views, negotiating and finding compromises, imagining reactions to emotions, and understanding and controlling self emotions are not influenced by the sentiment of the dyad's conversations. A possible explanation for this outcome is the minimal use of communication in these specific variables of the dyad's interaction. Although the dyad members communicated with each other throughout the duration of the task, the sentiment of the conversation was not a relatively important factor influencing these variables, However, these variables account for 3 out of the 8 collaboration factors, and therefore, considerably influence the team performance. To gain a complete understanding of the factors influencing team performance, it is essential for future studies to discover and examine other dynamics of a team's interaction during a collaborative task. However, these variables account for 3 out of the 8 collaboration factors, which is a considerably large portion of the collaboration factors.

From the comparison, it was found that the variance explained by the predictors was quite similar for Emotion Management and Co-construction, even though the latter was not predicted significantly. An explanation for the possibly lower adjusted R-squared value for the Co-construction model could be the predictors other than the mean polarity score being of less value. A solution for this could be the use of the backward entry method for the regression models to see if the results are improved. The results for this part of the study could have been influenced due to some other limitations as well. First of all, the sentiment analysis approach selected for this study could potentially be improved. Although, there are fewer resources available for the French language and conversations in general, a machine learning model trained on conversation environments like Twitter to test the difference in results. Secondly, it is possible that the number of positive and negative texts could be replaced with more appropriate measures reflecting variation in sentiments of individual texts. Since the transcripts were not recorded for all dyads, collaboration information was excluded for 11 dyads, which might have led to less reliable results than expected.

The second goal of this study was to examine the relationship between the emotions perceived by the participants at the end of the collaboration and the two different physiological signals: electrodermal activity (EDA) and heart rate variability (HRV). Based on a circumplex model of emotion, four different kinds of emotions (written in French) were selected: Ravi, Lasse, Detendu, and Frustre. Three different multiple linear regression models were generated to predict HRV and EDA from the selected emotions. Two of the three models were for the prediction of HRV, using two different measures as the outcome variable in each model (RMSSD and beats per minute). The third model was for the prediction of EDA, using EDA peaks per minute, also called skin conductance response (SCR) peaks, as the outcome variable. The results showed no significant effects observed in the regression models. Therefore, the coefficients from the regression models were visualized (using dots-and-whisker plots) to examine the effects of each emotion on the different measures of physiological signals.

The results showed Frustre and Detendu being negatively related to RMSSD while Ravi and Lasse were seen to be positively associated. The strongest association among all predictors was observed for Ravi, followed by Detendu. This particular result allows for the interpretation that emotions with positive valence exhibit a stronger association with RMSSD. On the other hand, all emotions except Detendu indicated a negative relation to beats per minute. The strongest predictor for this measure was Lasse, a low-arousal, negative-valence emotion. Furthermore, for the prediction of EDA peaks per minute, Ravi and Lasse were found to be negatively related, whereas Frustre and Detendu were positively related. The strongest effect was shown by Lasse, followed shortly by Detendu (both being emotions with low arousal). It can be understood from these results that emotions classified as being low-arousal are expected to have a relatively stronger relation with the number of EDA peaks per minute.

The first part of this study contributes to the field of research by improving understanding of the factors that influence team performance in dyads. This is because this study investigated a rather different variable as the driver of change in team dynamics. Moreover, this study helps understand which specific aspects of a team's performance are influenced by the sentiment of their conversations, and to what extent can the impact be seen. The second part of this study contributes by improving the current knowledge and frameworks within this domain. As discussed in Section 2.1 and 2.5, the current emotion recognition model for collaboration between dyads does not use physiological data as a possible indication of emotion. The results from the second part of this study showed that for two out of the three measures, patterns were observed as to what kind of emotion could exhibit stronger relations. From this, it can be understood that with further research, emotion recognition in this context stands to gain from different measures of physiological data. From this evaluation, this study adds to the findings of Chanel et al. (2017).

6 CONCLUSION

This study was primarily aimed at understanding the impact of conversation sentiment on subjective team performance in dyads. The research question posed to facilitate this goal was: **"Does positive conversation sentiment during a collaborative task predict subjective team performance in dyads?"**. The hypothesis generated for this question was: **"Higher (more positive) sentiment score leads to better team performance in a collaborative task"**. The results showed that conversation sentiment prediction two out of the eight collaboration factors significantly. Mean polarity score of the conversation showed a significant positive effect with Convergence, due to which the hypothesis made could not be rejected. Other than mean polarity score, the number of positive and the number of negative texts also showed a relatively strong, but not significant effect, with two different collaboration factors.

This study also aimed at understanding the relation between physiological signals recorded during a collaborative task and the emotions felt by the participants following the interaction. To this end, the following research question was posed: **"How does interpersonal physiological data relate to perceived emotions reported by the participants after a collaborative task in dyads?"**. The results did not show any significant relationship, however, interesting patterns were found. It was learned that emotions with low arousal scores are expected to show a stronger effect on EDA peaks per minute, and emotions with positive valence are expected to show a stronger effect on the Root Mean Square of Successive Differences between heartbeats.

Since conversation sentiment exhibited significant effects on team performance in dyads, further research with a larger dataset and more representative measures of a conversation's sentiment could help understand how to moderate the interactions to some extent for maximum performance. Furthermore, results from this study allow for a better understanding of computer-supported collaboration processes (scope of the EATMINT dataset). Future research in this area can focus on investigating different measures such as peak amplitude and peak onset for EDA, Standard deviation of RR intervals, or different frequency-domain measures for HRV. Three out of the eight collaboration factors showed relatively small effects with the predictor variables. Therefore, to gain a deeper understanding of the factors influencing team performance, it is essential for future studies to discover and examine other dynamics of a team's interaction during a collaborative task.

REFERENCES

Abirami, A., & Gayathri, V. (2017). A survey on sentiment analysis methods and approach. In *2016 eighth international conference on advanced computing (icoac)* (pp. 72–76).

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011). Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (lsm 2011)* (pp. 30–38).

Ahonen, L., Cowley, B., Torniainen, J., Ukkonen, A., Vihavainen, A., & Puolamäki, K. (2016). Cognitive collaboration found in cardiac physiology: Study in classroom environment. *PloS one*, *11*(7), e0159178.

Ahonen, L., Cowley, B. U., Hellas, A., & Puolamäki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: Eda and ecg study of pair-programming in a classroom environment. *Scientific reports*, *8*(1), 1–16.

Appelhans, B. M., & Luecken, L. J. (2006). Heart rate variability as an index of regulated emotional responding. *Review of general psychology*, *10*(3), 229–240.

Bestelmeyer, P. E., Kotz, S. A., & Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social cognitive and affective neuroscience*, *12*(8), 1351–1358.

Bilakhia, S., Petridis, S., Nijholt, A., & Pantic, M. (2015). The mahnob mimicry database: A database of naturalistic human interactions. *Pattern recognition letters*, *66*, 52–61.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".

Boiger, M., & Mesquita, B. (2012). The construction of emotion in interactions, relationships, and cultures. *Emotion review*, *4*(3), 221–229.

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (eda) & skin conductance responses (scrs) for psychological experiments. *Psychophysiology*, *49*(1), 1017–1034.

Caruelle, D., Gustafsson, A., Shams, P., & Lervik-Olsen, L. (2019). The use of electrodermal activity (eda) measurement to understand consumer emotions–a literature review and a call for action. *Journal of Business Research*, *104*, 146–160.

Chanel, G., Avry, S., Molinari, G., Bétrancourt, M., & Pun, T. (2017). Multiple users' emotion recognition: Improving performance by joint modeling of affective reactions. In *2017 seventh international conference on affective computing and intelligent interaction (acii)* (p. 92-97). doi: 10.1109/ACII.2017.8273584

Chanel, G., Bétrancourt, M., Pun, T., Cereghetti, D., & Molinari, G. (2013).

Assessment of computer-supported collaborative processes using interpersonal physiological and eye-movement coupling. In *2013 humaine association conference on affective computing and intelligent interaction* (pp. 116–122).

Chanel, G., & Mühl, C. (2015). Connecting brains and bodies: applying physiological computing to support social interaction. *Interacting with Computers*, *27*(5), 534–550.

Chartrand, T. L., van Baaren, R. B., & Bargh, J. A. (2006). Linking automatic evaluation to mood and information processing style: consequences for experienced affect, impression formation, and stereotyping. *Journal of Experimental Psychology: General*, *135*(1), 70.

Critchley, H. D. (2002). Electrodermal responses: what happens in the brain. *The Neuroscientist*, *8*(2), 132–142.

Elkins, A. N., Muth, E. R., Hoover, A. W., Walker, A. D., Carpenter, T. L., & Switzer, F. S. (2009). Physiological compliance and team performance. *Applied ergonomics*, *40*(6), 997–1003.

Forgas, J. P., & George, J. M. (2001). Affective influences on judgments and behavior in organizations: An information processing perspective. *Organizational behavior and human decision processes*, *86*(1), 3–34.

Goldstein, I. L., & Ford, K. (1993). *Training in organizations: Needs assessment, development, and evaluation . pacific grove, ca: Brooks.* Cole Publishing Company.

Gorman, J. C., Dunbar, T. A., Grimm, D., & Gipson, C. L. (2017). Understanding and modeling teams as dynamical systems. *Frontiers in psychology*, *8*, 1053.

Gujjar, J. P., & Kumar, H. P. (2021). Sentiment analysis: Textblob for decision making. *International Journal of Scientific Research & Engineering Trends*(7), 1097–1099.

Guzzo, R. A., & Dickson, M. W. (1996). Teams in organizations: Recent research on performance and effectiveness. *Annual review of psychology*, *47*(1), 307–338.

Henning, R., Armstead, A., & Ferris, J. (2009). Social psychophysiological compliance in a four-person research team. *Applied ergonomics*, *40*(6), 1004–1010.

Henning, R. A., Boucsein, W., & Gil, M. C. (2001). Social–physiological compliance as a determinant of team performance. *International Journal of Psychophysiology*, *40*(3), 221–232.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, *9*(3), 90–95.

Kazi, S., Khaleghzadegan, S., Dinh, J. V., Shelhamer, M. J., Sapirstein, A., Goeddel, L. A., . . . Rosen, M. A. (2021). Team physiological dynamics: A critical review. *Human factors*, *63*(1), 32–65.

Kemp, A. H., & Quintana, D. S. (2013). The relationship between mental and physical health: insights from the study of heart rate variability. *International Journal of Psychophysiology*, *89*(3), 288–296.

Kotzé, E., & Senekal, B. (2018). Employing sentiment analysis for gauging perceptions of minorities in multicultural societies: An analysis of twitter feeds on the afrikaner community of orania in south africa. *TD: The Journal for Transdisciplinary Research in Southern Africa*, *14*(1), 1–11.

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition*, *112*(3), 473–481.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological psychology*, *84*(3), 394–421.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1–167.

Liu, B., et al. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, *2*(2010), 627–666.

Liu, P., Joty, S., & Meng, H. (2015). Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1433–1443).

Liu, Z., Xu, A., Guo, Y., Mahmud, J. U., Liu, H., & Akkiraju, R. (2018). Seemo: A computational approach to see emotions. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1–12).

Loria, S. (2018). textblob documentation. *Release 0.15*, *2*.

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., ... Chen, S. (2021). Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior research methods*, *53*(4), 1689–1696.

Mather, M., & Thayer, J. F. (2018). How heart rate variability affects emotion regulation brain networks. *Current opinion in behavioral sciences*, *19*, 98–104.

McKinney, W., et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th python in science conference* (Vol. 445, pp. 51–56).

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093–1113.

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, *2*(1), 63–86.

Molinari, G., & Avry, S. (2018). Flow in computer-supported collaborative problem-solving. International Society of the Learning Sciences, Inc.[ISLS].

Pan, J., & Tompkins, W. J. (1985). A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*(3), 230–236.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rhouati, A., Berrich, J., Belkasmi, M. G., & Bouchentouf, T. (2018). Sentiment analysis of french tweets based on subjective lexicon approach: Evaluation of the use of opennlp and corenlp tools. *J. Comput. Sci.*, *14*(6), 829–836.

RStudio Team. (2015). Rstudio: Integrated development environment for r [Computer software manual]. Boston, MA. Retrieved from http://www.rstudio.com/

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Saif, H., He, Y., & Alani, H. (2012). Semantic sentiment analysis of twitter. In *International semantic web conference* (pp. 508–524).

Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in psychology*, *5*, 1040.

Stuldreher, I. V., Thammasan, N., Van Erp, J. B., & Brouwer, A.-M. (2020). Physiological synchrony in eeg, electrodermal activity and heart rate detects attentionally relevant events in time. *Frontiers in Neuroscience*, *14*, 575521.

Van Gent, P., Farah, H., Van Nes, N., & Van Arem, B. (2019). Heartpy: A novel heart rate algorithm for the analysis of noisy signals. *Transportation research part F: traffic psychology and behaviour*, *66*, 368–378.

Van Kleef, G. A., Homan, A. C., & Cheshin, A. (2012). Emotional influence at work: Take it easi. *Organizational Psychology Review*, *2*(4), 311–339.

Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. Scotts Valley, CA: CreateSpace.

Walker, A. D., Muth, E. R., Switzer III, F. S., & Rosopa, P. J. (2013). Predicting team performance in a dynamic environment: A team psychophysiological approach to measuring cognitive readiness. *Journal of cognitive engineering and decision making*, *7*(1), 69–82.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the acl 2012 system demonstrations* (pp. 115–120).

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and instruction*, *17*(4), 416–426.

Yost, C. A., & Tucker, M. L. (2000). Are effective teams more emotionally intelligent? confirming the importance of effective communication in teams. *Delta Pi Epsilon Journal*, *42*(2), 101.

APPENDIX A

All the additional tables are provided here.

Table 3: Standardized beta coefficients for the regression models of each collaboration factor.

Collaboration Factor	Mean Polarity	Number of Positive Texts	Number of Negative Texts
Grounding & Coordination	-0.09	0.10	0.40
Degree of Convergence	0.77	0.05	0.18
Emotion Modeling	0.38	-0.20	0.23
Transactivity	-0.54	0.37	-0.07
Emotion Management	-0.79	0.55	-0.20
Degree of Conflict	-0.41	-0.08	0.19
Confrontation & Consensus Building	-0.47	-0.08	-0.18
Co-Construction	-0.21	0.12	0.50

Table 4: Adjusted R-squared values and percentage of variance explained in the data for the regression models of each collaboration factor.

Collaboration Factor	Adjusted R-Squared	Percentage of Variance Explained in the Data
Grounding & Coordination	0.087	8.7
Degree of Convergence	0.504	50.4
Emotion Modeling	-0.103	-
Transactivity	0.087	8.7
Emotion Management	0.317	31.7
Degree of Conflict	0.106	10.6
Confrontation & Consensus Building	0.115	11.5
Co-Construction	0.306	30.6

Table 5: The coefficients of the regression models for research question 2, with outcome variables as rows and predictor variables as columns.

Variable	Ravi	Frustre	Lasse	Detendu
RMSSD	8.68	-6.21	5.74	-7.53
BPM	-0.65	-0.04	-1.32	0.20
EDA peaks per minute	-0.04	0.06	-0.20	0.18

Table 6: Standardized beta coefficients of the regression models for research question 2, with outcome variables as rows and predictor variables as columns.

Variable	Ravi	Frustre	Lasse	Detendu
RMSSD	0.25	-0.19	0.15	-0.28
BPM	-0.10	-0.01	-0.17	0.04
EDA peaks per minute	-0.03	0.05	-0.13	0.17

Table 7: List of all the emotions in the questionnaire along with their English translation obtained using the DeepL translator.

Emotion	Translation in English
Ravi	Delighted
Concentre	Concentration
Interesse	Interested
Satisfait	Satisfied
Empathique	Empathy
Confiant	Confident
Amuse	Amused
Detendu	Relaxed
Reconnaissant	Acknowledging
Soulage	Relief
Stresse	Stress
Enerve	Angry
Surpris	Surprised
Irrite	Irritated
Envieu	Envious
Anxieu	Anxious
Insatisfait	Unsatisfied
Confus	Confused
Frustre	Frustrated
Lasse	Tired

APPENDIX B

All the additional figures are provided here.

Figure 12: The formula used for the calculation of Root Mean Square of Successive Difference between heart beats (RMSSD). Source: iMotions

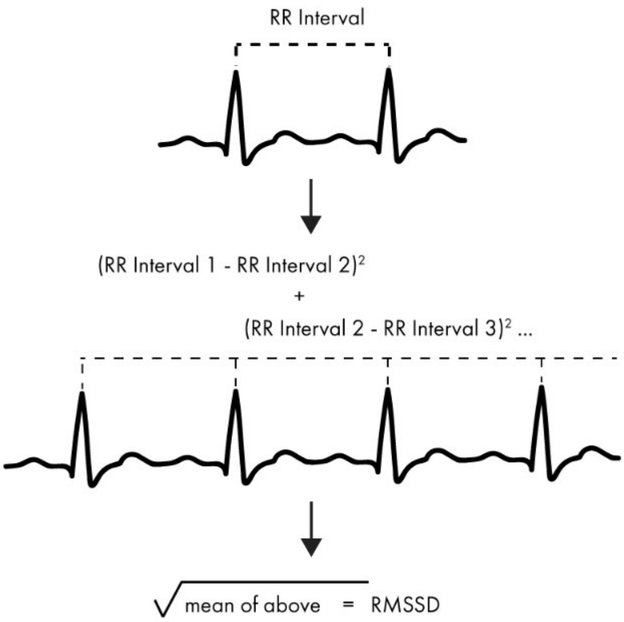


Figure 13: Distribution of RR intervals shown from a random participant's derived HRV data.

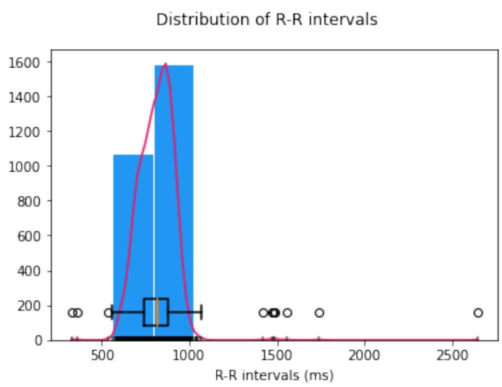


Figure 14: Raw ECG signal for a randomly selected participant.

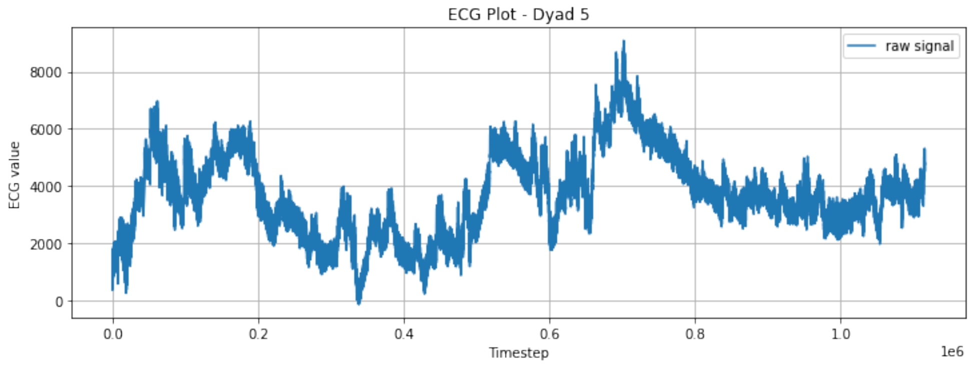


Figure 15: Zoomed-in view of peak detection on the EDA signal.

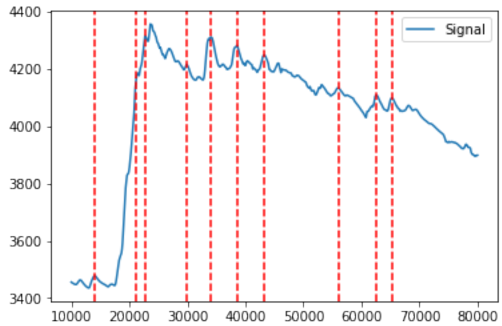


Figure 16: Erroneous EDA signals for Participant 1 from Dyad 4 (this dyad was removed from the analysis).

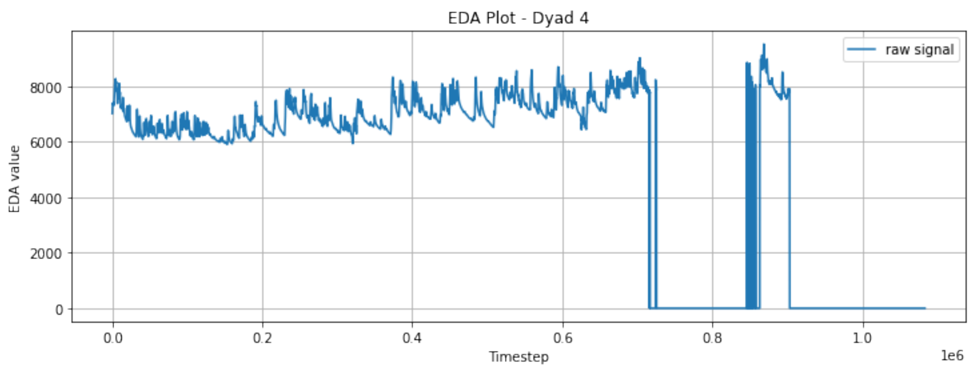


Figure 17: Erroneous ECG signals for Participant 1 from Dyad 29 (this dyad was removed from the analysis).

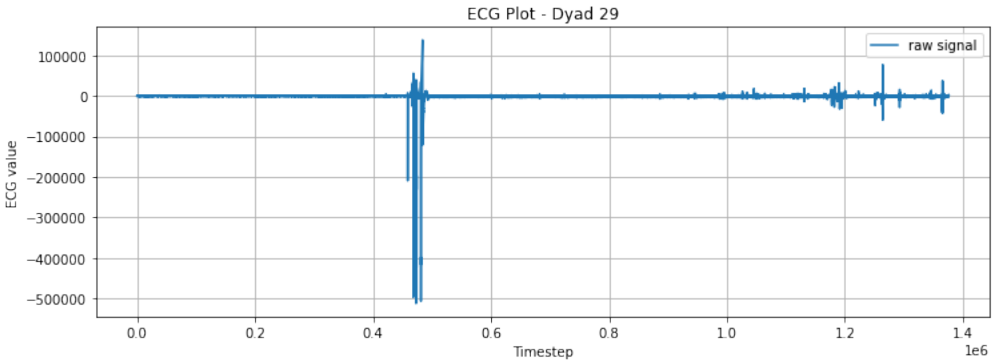


Figure 18: Erroneous ECG signals for Participant 1 from Dyad 4 (this dyad was removed from the analysis).