Predicting Subjective Team Performance Using Multimodal, Single-Modality and Segmented Physiological Data

Frank Arts
STUDENT NUMBER: 2018932

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

Thesis committee:
Supervisor: Travis J. Wiltshire
Second reader: Emmanuel Keuleers

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science & Artificial Intelligence
Tilburg, The Netherlands
June 2020

Preface

I would like to thank my supervisor, Travis, for his continued involvement and support, both during my thesis project and during my position as a research assistant. Without his support I would not even have known that this field of research exists, and I definitely would not have been able to understand the methods of analysis in this field of research. Due to the Corona regulations, I experienced difficulties both in my thesis and in my job as a research assistant. Travis went out of his way to offer me alternative work, and I am incredibly grateful for that. Without this support, I would not have been able to focus on my thesis and probably would not have been able to complete it this semester.

I would also like to thank my thesis group members, because their involvement in the thesis project made the entire process a lot more enjoyable.

Finally I would like to thank my girlfriend Benthe for her continued support, both during the thesis project and in general, and for making sure that I could keep working on my thesis from home, in a comfortable working environment.

Predicting Subjective Team Performance using Multimodal, Single-Modality and Segmented Physiological Data

By Frank Arts

Abstract
In this study, multimodal team coordination measures were compared to single-modality team coordination measures to determine how well they can predict subjective team performance. To this end, multimodal and single-modality team coordination measures were calculated from heart rate variability data and electrodermal activity data found in the EATMINT dataset (Chanel et al., 2013), using Cross Recurrence Quantification Analysis and Multidimensional Recurrence Quantification Analysis. Multiple linear regression models were generated using both multimodal and single-modality team coordination measurements as predictors for eight different subjective team performance measures. By comparing the standardized beta coefficients of the different team coordination measures, it was found that multimodal coordination measures were more important for predicting three out of the eight subjective team performance measures. For two out of the eight subjective team performance measures, only one of the multimodal team coordination measures was more effective at predicting the subjective team performance measure than its single-modality counterpart.

Additionally, this study investigates how segments of physiological data can be used for predicting subjective team performance. By segmenting the physiological data into windows in different ways, the effect of window size and window location on the ability to predict subjective team performance was investigated. It was found that some windows were more critical for predicting subjective team performance than others, and that smaller window sizes generally allowed for more accurate predictions of subjective team performance.

1. Introduction

There are many issues that we face in society that are too complex to solve individually, either due to the complexity or due to the multidisciplinary nature of these issues. As a result, these issues require collaboration between different individuals. To ensure that such issues can be tackled efficiently, it is important to understand what makes team members work together effectively. To do so, a thorough understanding of the underlying mechanisms that play a role in team dynamics is required.

A clear understanding of the different underlying mechanisms that play a role in team dynamics could allow for predicting the effectiveness of specific teams to some extent. The ability to predict how well a certain team is likely to perform would be invaluable for society. Teams that are identified as being likely to underperform could be subjected to intervention measures in an attempt to improve their expected performance. For example, internal team composition could be changed in an attempt to improve the expected performance.

If the prediction of team performance is able to distinguish between different aspects of team performance, more suitable intervention measures could be applied as needed to improve team performance more efficiently. For example, if a lack of communication is identified early on, a working environment focusing on improving communication could be introduced to supplement the lack of communication. Likewise, if a lack of emotional understanding within a team is identified to likely negatively impact the team performance, measures could be taken to improve emotional understanding amongst team members. Maximizing expected team performance in such a way could save a lot of time and money, but would require a thorough understanding of underlying mechanisms in team dynamics.

In this thesis, teams will be investigated from a dynamical systems theory perspective, as described by Gorman et al. (2017). They argue that team dynamics should be studied through the interactions between team members rather than focusing on the behavior of individual team members. According to the dynamical systems theory, synchronization occurs when two or more individual components in a dynamical system interact with each other. This phenomenon can also be applied to teams, where team members have an effect on each other as they become informationally coupled.

One modality in which such team coordination effects have been observed is through physiological signals. Studies have found that different types of physiological signals synchronize to some extent between team members as they work together. For example, physiological synchrony has been observed in heart activity, skin conductance and even EEG signals.

The link between physiological team coordination and team performance has been investigated by many studies. However, the majority of studies investigating physiological team coordination have only investigated physiological coordination in dyads (Kazi et al., 2019), and they often used only one physiological measurement. Even amongst the studies using teams of three members or more or using multiple types of physiological measurements, many of them used methods that are not suitable for processing such multidimensional data. Instead, they split up teams into dyads for the actual data analysis, or they analyze the different physiological measurements separately. As a result, findings from these studies might be lacking important information, as patterns that emerge at higher levels are not discovered when only analyzing data pairwise.

Research Questions

The present study aims to determine whether patterns embedded across multiple modalities might be important for understanding team dynamics, and thus whether it is necessary to use analysis techniques that are capable of adequately dealing with multidimensional data. To do so, this study aims to investigate how important multimodal team coordination measures are when predicting subjective team performance, compared to single-modality team coordination measures. To this end, the following research question was generated:

RQ1: How well can multimodal team coordination measures predict subjective team performance compared to single-modality team coordination measures?

In addition to this, the present study also aims to investigate how well subjective team performance could be predicted based on only a part of the physiological data collected during collaborative tasks. Being able to predict subjective team performance based on a small segment of physiological data would be useful for society, as problems in teamwork could be identified earlier on. This would in turn allow for earlier interventions, potentially avoiding bad team performance.

From an academical point of view, predicting subjective team performance using only a segment of physiological data could improve our understanding of underlying mechanisms of subjective team performance. If subjective team performance can be predicted significantly based only on a small segment of data, then that segment must be of critical importance subjective team performance. Otherwise, the lack of complete data would make it impossible to predict subjective team performance. As such, the following research questions were generated:

RQ2: Can segments of physiological data be useful for predicting subjective team performance?

RQ2a: How does the size of segments of physiological data affect the usefulness for predicting subjective team performance?

RQ2b: Which segments of physiological data are more crucial for predicting subjective team performance than others?

2. Related Work

This study will use team coordination measures calculated from heart rate variability and electrodermal activity data to predict subjective team performance. The reason for using these physiological signals is that heart rate variability coordination has already been shown to correlate with team performance (Henning et al. 2001). Additionally, studies have indicated that electrodermal activity coordination is related to team performance as well. Ahonen et al. (2018) found that a higher skin conductance synchrony at the start of a task was linked to higher team performance, while skin conductance synchrony at the end of a task was higher for teams that failed their tasks

A number of studies have found a correlation between physiological coordination and team performance already (Henning et al., 2001, Walker et al. 2013, Ahonen et al. 2016, Tschacher & Meier, 2019), but the majority of studies on physiological team coordination have focused on dyads (Kazi et al., 2019). Even when studies investigate team coordination in teams of three or more members, they often split the groups up into dyads for the actual analysis. For example, Elkins et al. (2009) collected data from teams of four members, but due to movement artifacts they chose to only look at the highest quality data of two of the four members. Henning, Armstead and Ferris (2009) also investigated teams of four participants, but chose to calculate team coordination pairwise by splitting up the group into dyads for the analysis, rather than calculating multivariate team coordination.

Additionally, many studies only use one type of physiological measurement in their coordination measures (Kazi et al., 2019), or when they use multiple physiological measurements they do not combine them into a multimodal coordination measure (Henning et al., 2001; Henning et al., 2009; Ahonen et al., 2018). As a result, the team coordination measurements calculated in those studies do not capture the patterns in the data that is embedded across multiple modalities.

While only a few studies have looked at physiological coordination in triads or larger groups or across more than one modality, there are some studies that have investigated suitable methods to measure some type of coordination using multivariate data. For example, Richardson et al. (2012) created the cluster-phase method for assessing movement synchrony in groups of six, and Wallot et al. (2016) created and demonstrated the MdRQA tool for analyzing skin conductance coordination in triads.

Additionally, Eloy et al. (2019) used MdRQA for assessing speech rate, body movement and galvanic skin response coordination in triads on a multimodal level. More specifically, they investigated how recurrence across multiple modalities could predict team performance, and they found that multimodal recurrence was predictive of the valence of collaboration.

A search of the literature revealed no studies that compared multimodal coordination measures to single-modality coordination measures when predicting team performance. However, Fusaroli and Tylén (2016) investigated how two different approaches differ in their effectiveness for predicting collective performance of dyads in a collaborative task. While the approaches tested in this research are different from the ones in the current study in that they pertain to dialog rather than physiological coordination, the statistical analyses are suitable for comparing utility of different approaches such as multimodal coordination and single-modality coordination measures.

The present study investigates how important multimodal patterns are for predicting subjective team performance by answering RQ1. It is important to identify how important multimodal coordination measures are compared to single-modality measures, so that future researchers can decide which measures they should incorporate into their studies. In the present study this is investigated by comparing single-modality team coordination measures extracted solely from heart rate variability (HRV), or solely from electrodermal activity (EDA) to multimodal team coordination measures calculated from the combined HRV and EDA signals.

In order to effectively compare multimodal team coordination measures to single-modality team coordination measures, an analysis technique is needed that can calculate multimodal team coordination measures in a similar way that other analysis techniques calculate single-modality team coordination measures. For this reason, the Multidimensional Recurrence Quantification Analysis (MdRQA) and Cross Recurrence Quantification Analysis (CRQA) methods were used as described by Wallot and Leonardi (2018).

These two analyses are both Recurrence Quantification Analyses, with MdRQA being capable of calculating team coordination at a multimodal level, and CRQA being very suitable for calculating team coordination at single-modality level. Since both analyses are derived from Recurrence Quantification Analysis, they also both calculate similar team coordination measurements. An explanation of how these two analyses work is given in section 3.3.

MdRQA can also be used for calculating team coordination in teams with more than 2 members, making it an important new method for the field of team dynamics. While the current study uses data from dyads, hopefully future studies will be able to reproduce the results of this study using triads or larger teams without needing to change the analysis techniques.

Aside from comparing multimodal team coordination measures to single-modality team coordination measures, the present study also aims to investigate the effectiveness of segments of physiological data in predicting subjective team performance. There has been limited research into predicting team performance based on smaller segments of data. Henning and Korbelak (2005) used physiological coordination of teams to predict the team performance of future tasks. They found that higher physiological coordination during a task period before a change in control dynamics was made was correlated to a lower team tracking error after a change in control dynamics was made, thus improving team performance.

Schoenherr et al. (2019) investigated how nonverbal (movement) coordination between patient and therapist dyads can be used to predict premature termination of psychotherapy for social anxiety disorder. Premature termination of psychotherapy sessions could be seen as an inverse team performance measure, and movement synchrony is very similar in data to physiological synchrony.

Additionally, Carrère and Gottman (1999) used small segments of conversational data to predict whether couples would stay together or divorce over a 6-year timespan. They found that patterns embedded in the first 3 minutes of quantitative affect data were already enough to predict this outcome.

Since little research has been done into the use of segments of data for predicting team performance, the results of this study will be able to contribute to the understanding of important team dynamics patterns located in partial segments of data. Additionally, the study will offer a method for locating critical segments in any combination of different types of time-scale data for predicting subjective team performance.

3. Method

3.1 Dataset

This work uses the data from the publicly available EATMINT dataset, as described by Chanel et al. (2013). This dataset consists of multimodal data recordings of 30 same gender dyads as they collaborate remotely. The dyads were asked to use the collaborative environment DREW to design a slogan against violence in school in less than 45 minutes. During the first part of the experiment, teams were instructed to generate as many slogan ideas as possible. During the second part of the experiment, teams were instructed to debate the relevance of their slogan ideas, and to suppress the less relevant ideas. For the final part of the experiment teams were instructed to find a consensus on the best slogan. During this entire experiment, the team members were able to communicate using the DREW environment and orally using headsets, but they were not able to see each other.

The data provided by this dataset consists of raw unfiltered ECG and EDA signals, as well as the responses to a self-report questionnaire on the participants' emotions and perceived collaboration. Additionally, the dataset contains eight factors that describe different aspects of the perceived collaboration for each participant: Grounding & Coordination, Degree of Conflict, Degree of Convergence, Confrontation & Consensus Building, Co-construction, Emotion Management, Emotion Modeling and Transactivity. These eight factors were the result of a factorial analysis ran by Chanel et al. (2013) on the self-report questionnaire. The present study will be using these eight collaboration factors as different subjective team performance measures. In the rest of this paper, these factors will be referred to as 'Grounding', 'Conflict', 'Convergence', 'Confrontation', 'Co-construction', 'Emotion Management', 'Emotion Modeling' and 'Transactivity'.

3.2 Physiological Data

Heart rate variability

Heart rate variability was calculated from the raw ECG data. The R peaks were extracted from the raw ECG signal using the Pan & Tompkins algorithm as implemented by the rsleep package (Bouchequet, 2020) in R (R Core Team, 2020), using a band-pass filter between 0.05 Hz and 40 Hz as suggested by Chanel et al. (2013). The R peaks time-series were then converted into an Inter-Beat Interval (IBI), by splitting the data up into 2-second consecutive and non-overlapping windows and calculating the average IBI for each of the 2 second windows.

The resulting IBI time-series had a sampling rate of 0.5Hz. This specific sampling rate was chosen to avoid sampling the same IBI value in two consecutive windows, as a result of no new R-peaks appearing in one of the windows. Having repeated values in the IBI time-series as a result of inadequate sampling rate could interfere with the integrity of the recurrence analysis. However, with this sampling rate this could only occur when a participant's heart rate rises above 120 bpm, which is rare during the type of task that was performed for this experiment.

Electrodermal Activity

In order to generate multimodal measures using the MdRQA method, the different physiological signals need to be sampled at the same sampling rate. For this reason, EDA was also sampled at 0.5 Hz. To be more specific, EDA was sampled once every 2 seconds, in the exact center of each 2-second window that the IBI intervals were calculated for. This EDA value was then smoothed by taking the average of the surrounding 0.5s window. This resulted in an EDA time-series sampled at 0.5Hz, with exactly the same amount of data points as the IBI time-series.

As a result of missing data or invalid data, in either the EDA or the ECG data from either member of a dyad, dyads 4, 27 and 29 were excluded from the analysis. As a result, only 26 dyads were used for the final analyses.

3.3 Team coordination measures

In this study, team coordination was evaluated on a multimodal level, in addition to single modality levels. For this reason, both Multidimensional Recurrence Quantification Analysis (MdRQA) and Cross Recurrence Quantification Analysis (CRQA) were used. These two analyses are both recurrence-based analyses, meaning that they both quantify how often the same states reoccur in a dynamical system. MdRQA was used for assessing team coordination on the multimodal level, while CRQA was used for assessing team coordination within a single modality. Both of these recurrence-based analyses make use of recurrence plots, allowing for the same team coordination measures to be extracted from the two analyses.

Multidimensional Recurrence Quantification Analysis

Multidimensional Recurrence Quantification Analysis (MdRQA) is a recurrence-based analysis that quantifies how often recurrence patterns occur within a set of two or more time-series. MdRQA allows for estimating coordination across multiple time-series, making it a very suitable method for evaluating team coordination in groups of three or more participants, or for evaluating coordination across multiple modalities.

MdRQA takes multiple time-series of the same length as input, and compares the state that the multiple time-series are in at one point in time to the state that the multiple time-series are in at a different point of time. It calculates the distance between two states pairwise to see how similar the states are. If the distance between the two states is less than a manually determined radius threshold, than the combination of states is marked as a recurrent point. The result of this process is a recurrence plot, where the combined time-series is plotted against itself, and any combination of two states that is similar to each other is marked with a recurrent point.

An example of such a recurrence plot can be found in figure 1a. The black cells in this image represent the recurrent points, while the white cells represent non-recurrent points. Since the MdRQA recurrence plot compares a combined time-series to itself, the diagonal line of identity always consists of recurrent points, as a state is always identical to itself. A recurrent point close to the line of identity indicates that two states were similar to each other with a short lag in between, whereas a recurrent point further away from the line of identity indicates that two states were similar to each other with a longer lag in between. Finally, it should be noted that the MdRQA recurrence plot is mirrored in the line of identity, as any combination of two states that is found on one side of the line of identity is also found on the other side of the line of identity, with the same distance measure.

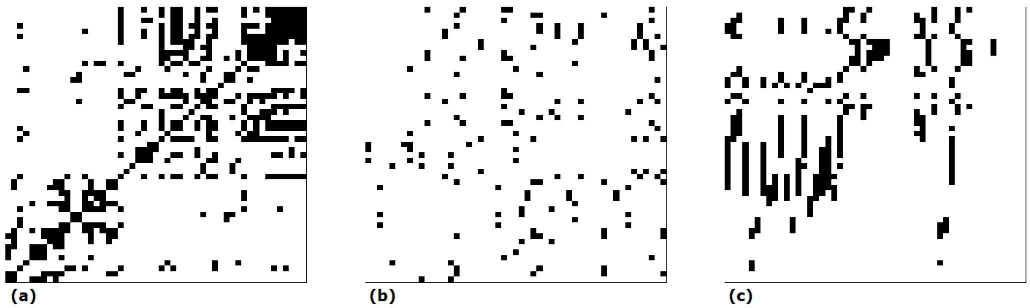
Cross Recurrence Quantification Analysis

Like MdRQA, Cross Recurrence Quantification Analysis (CRQA) is also a recurrence-based analysis. However, CRQA focuses on comparing only two timeseries. Instead of grouping all time-series together and comparing that combined time-series to itself, CRQA compares two separate time-series to each other. It compares each data point in one time-series to each data point in the other time-series. If the distance between the two values is larger than a manually selected radius threshold, than the combination of these two data points is marked as a recurrent point.

An example of a partial CRQA recurrence plot can be found in figure 1b and 1c. Each black cell represents a moment in time where the value of one of the time-series was very similar to the value of the other time-series. Once again, the recurrent points closer to the center diagonal represent recurrence at smaller lag, while recurrent points further away from the center diagonal represent recurrence at greater lag. Unlike the MdRQA-generated recurrence plot, the CRQA recurrence plot does not have a line of identity. The reason for this is that two different time-series are compared, so the values from the two time-series at any given time do not have to match. Additionally, the CRQA recurrence plot is not symmetrical.

Figure 1.

Recurrence plots generated by **(a)** MdRQA using IBI and EDA signals, **(b)** CRQA using EDA signals and **(c)** CRQA using IBI signals. All three recurrence plots were generated using the same 100s segment of data.



Team coordination measures

The recurrence plots generated by CRQA and MdRQA can be used to extract different team coordination measures. One common measure extracted from recurrence plots is the recurrence rate. Recurrence rate describes the percentage of points within a recurrent plot that are marked as recurrent points. Recurrence rate can be used as an estimate for team-level regularity, with high recurrence rate indicating that a lot of states have previously occurred in the set of time-series, thus indicating high team-level regularity. This measurement does not look at further nuances such as the lag between the first and second state, or the context of how close other recurrent points are. Recurrence rate can be calculated by dividing the total number of recurrent points by the size of the recurrence plot.

Another measure that can be extracted from recurrence plots is determinism, which describes the percentage of recurrent points that are diagonally adjacent to other recurrent points. Determinism focuses on sequences of recurrence rather than individual recurrent points, as diagonally adjacent points in the recurrence plot directly follow each other chronologically. This means that the determinism measure is higher when there are more continuous sequences of recurrence. Determinism is calculated by dividing the number of diagonally adjacent recurrent points by the total number of recurrent points in the recurrence plot.

Recurrence rate (%REC) and determinism (%DET) were measured on both multimodal and single-modality levels. Multimodal %REC and %DET were extracted from the recurrence plots generated by the MdRQA function (Wallot & Leonardi, 2018), using z-score normalized IBI time-series from both team members of a dyad as well as the z-score normalized EDA time-series from both team members as input. For the single-modality measurements, %REC and %DET in the IBI modality were extracted using the CRQA function (Wallot & Leonardi, 2018) with only the z-score normalized IBI time-series of both members of a dyad as input. Additionally, the %REC and %DET in the EDA modality were measured using CRQA with only the z-score normalized EDA time-series of both members of a dyad as input. In total six different team coordination measures were extracted from each dyad.

To allow for comparing of team coordination measures between dyads, the analyses used to measure the team coordination measures need to use the same parameters. For this study, that means that a single radius threshold needed to be specified for all MdRQA analyses, another radius threshold needed to be specified for the CRQA analyses on the IBI level, and yet another radius threshold needed to be specified for the CRQA analyses on the EDA level. As recommended by Wallot and Leonardi (2018), the radius parameter was determined by adjusting it until the %REC for all dyads was between 1% and 5%. This resulted in a radius of 0.35 for the MdRQA method, a radius of 0.07 for the CRQA method on the IBI level, and a radius of 0.078 for the CRQA method on the EDA level.

To account for the differences in range for the different time-series, z-score normalization was used for all time-series. Without such normalization, the measures would have been based on differences in magnitude rather than sequential similarity (Wallot & Leonardi, 2018). Additionally, prior to calculating the %REC and %DET for the MdRQA recurrence plots, the line of identity was removed to avoid skewed measures.

Subjective Team Performance
The eight different collaboration factors resulting from the factorial analysis ran by Chanel et al. (2013) on the self-report questionnaire were used as subjective team performance measures. For each collaboration factor, the score for the two members of a dyad was averaged to obtain a dyad-level score. An explanation of each factor can be found in table 1.

Table 1
Factors used as a measure of subjective team performance (source: 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); participant's 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)

4. Results

The descriptives of team coordination measures extracted from the MdRQA and CRQA analyses can be found in table 2. The team-level collaboration factors, which function as subjective team performance measures, can also be found in table 2.

Table 2
Descriptives of the six team coordination measures and the eight collaboration factors

Team coordination measure	Mean	SD	Range
Multimodal %REC – IBI + EDA	3.29	0.58	2.26 – 4.82
Multimodal %DET – IBI + EDA	40.03	7.83	25.71 – 60.69
Single Modality %REC – IBI	4.12	0.22	3.89 – 4.75
Single Modality %DET – IBI	11.32	1.36	9.07 – 15.10
Single Modality %REC – EDA	4.42	0.28	3.84 – 4.94
Single Modality %DET – EDA	47.58	14.50	26.43 – 79.24
Collaboration Factor	**Mean**	**SD**	**Range**
Grounding & Coordination	-0.02	0.67	-1.46 – 1.39
Degree of Conflict	-0.01	0.69	-0.92 – 2.18
Degree of Convergence	-0.02	0.61	-1.56 – 1.16
Confrontation & Consensus building	0.07	0.62	-1.39 – 1.53
Co-Construction	-0.06	0.84	-1.35 – 1.51
Emotion Management	0.01	0.75	-1.1 – 1.61
Emotion Modeling	0.02	0.79	-1.28 – 2.02
Transactivity	0.03	0.71	-1.21 – 1.26

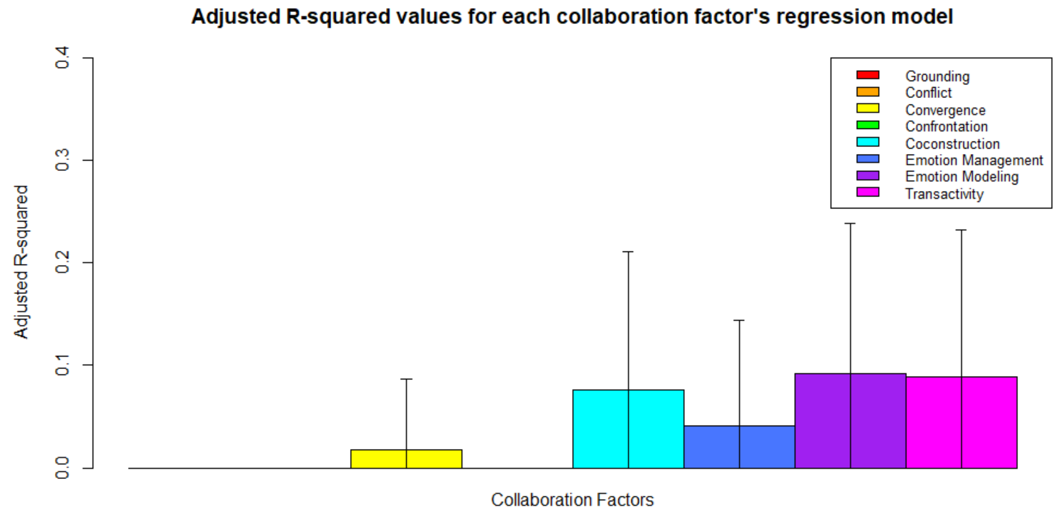
(RQ1) *How well can multimodal team coordination measures predict subjective team performance compared to single-modality team coordination measures?*

In order to compare multimodal team coordination measures to single-modality team coordination measures, multiple linear regression models were generated using both multimodal team coordination measures and single-modality team coordination measures as predictors. Since eight different collaboration factors represent subjective team performance, eight different multiple linear regression models were generated, one for each collaboration factor.

Each multiple linear regression model included six predictor variables: %REC and %DET at the multimodal level, %REC and %DET of the IBI signals and %REC and %DET of the EDA signals. These six predictors were checked for multicollinearity, but no predictors were removed as the strongest correlation between predictors was .72. The adjusted R^2 values and confidence intervals of the multiple linear regression models are visualized in Figure 2. While each of these regression models did not significantly predict the collaboration factors at a 95% confidence interval, internal comparisons of how strong the effect of each predictor is on the collaboration factor can be made.

Figure 2

The adjusted R^2 values and confidence intervals for the multiple linear regression models predicting each of the eight collaboration factors. 'Grounding', 'Conflict' and 'Confrontation' had an R-squared value of less than 0.



The effectiveness of the six predictors was examined by generating the standardized beta coefficients. The standardized beta coefficients allow for comparing how strong the effect of each team coordination measure is on the collaboration factor. Figure 3 shows the standardized beta coefficients for all six predictors for every collaboration factor. The exact values of each standardized beta coefficient can be found in Appendix A.

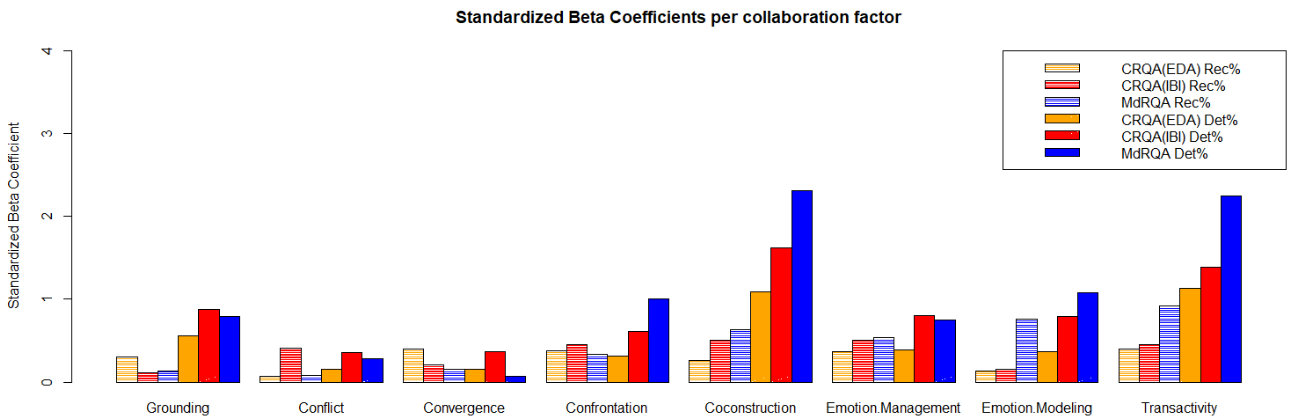
The results indicate that the multimodal team coordination measures have a stronger effect on 'Co-Construction', 'Emotion Modeling' and 'Transactivity' than the single-modality team coordination measures, as the multimodal standardized beta coefficients are larger.

For 'Emotion Management' and 'Confrontation', the results are mixed. In terms of recurrence rate, the multimodal measurement has a stronger effect on 'Emotion Management' than the single-modality measurements, while for determinism the single-modality measurements have a larger effect on 'Emotion Management' than the multimodal measurement. For 'Confrontation' the opposite is true, as the single-modality recurrence rate measurements have a larger effect on 'Confrontation' than the multimodal recurrence rate, while the multimodal determinism measure has a larger effect than the single-modality determinism measures.

Finally, for the factors 'Grounding', 'Conflict' and 'Convergence', the single-modality measurements have a stronger effect than the multimodal measurements.

Figure 3

Standardized beta coefficients of the single-modality and multimodal team coordination measures in eight different regression models; grouped by collaboration factor



(RQ2) Can segments of physiological data be useful for predicting subjective team performance? To investigate how useful segments of physiological data are for predicting subjective team performance, team coordination measures extracted from segments of physiological data were tested in multiple linear regression models to see how well they can predict subjective team performance. To this end, the existing recurrence plots were split up into non-overlapping windows along the diagonal line. An example of this can be seen in Figure 4. Team coordination measures were extracted from these segmented windows, and multiple linear regression models were generated based on those team coordination measures. In this way, the effect of segmented window size and segmented window location could be evaluated for different window sizes and different collaboration factors. This in turn allows for answering RQ2a and RQ2b as well.

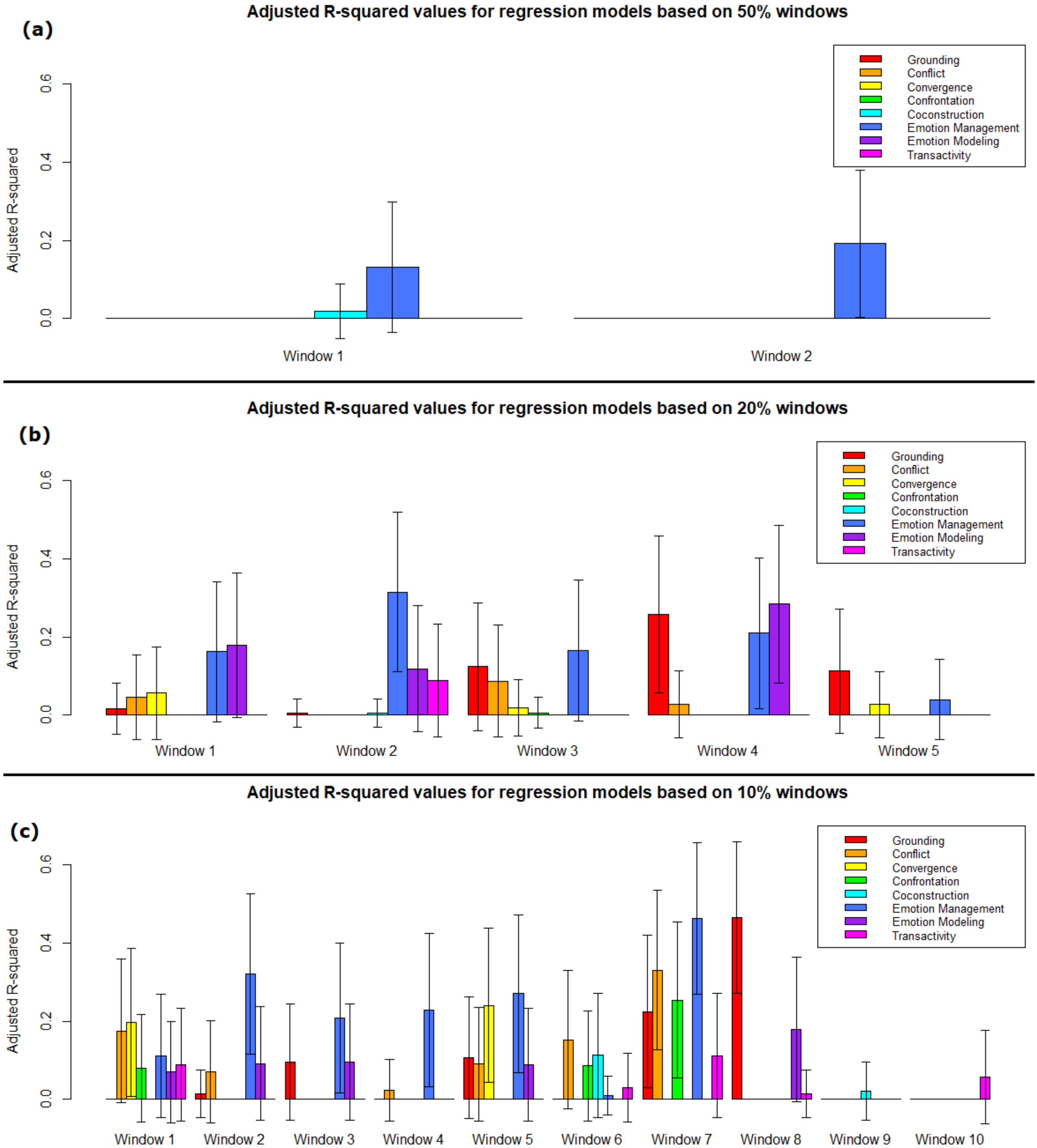
Figure 4

An example of how segmented windows can be extracted from the recurrence plot. This example shows the recurrence plot of the MdRQA analysis, with the red rectangles showing the segmented windows that are extracted. The window size in this example is 20%, resulting in a total of 5 windows.

Three different windowing sizes were used in this approach: A window size of 50%, resulting in two non-overlapping windows; a window size of 20%, resulting in five non-overlapping windows and a window size of 10%, resulting in ten non-overlapping windows. For every window, the six team coordination measures were calculated. Eight regression models were generated, one for every collaboration factor, using the six team coordination measures as predictors. The adjusted R^2 values for these regression models can be seen in Figure 5a for a window size of 50%, Figure 5b for a window size of 20% and Figure 5c for a window size of 10%.

Figure 5

The adjusted R^2 values and confidence intervals for the regression models predicting the eight different collaboration factors for each window. The window size is **(a)** 50%, **(b)** 20%, **(c)** 10%.



Window size 50%

For the window size of 50%, only the regression models predicting 'Co-construction' and 'Emotion Management' were able to predict team performance to some extent, as their adjusted R^2 values were larger than 0. However, only the regression model predicting 'Emotion Management' based on data in window 2 was able to significantly predict `Emotion Management`, as 0 was not in the confidence interval (*Adj.* $R^2 = 0.192$, $p < 0.05$).

Window size 20%

For the window size of 20%, more regression models were able to predict team performance to some extent, having adjusted R^2 values higher than 0. Both window 2 and window 4 produced some significant regression models. The team coordination measures based on window 2 were able to significantly predict 31.5% of the variability in the factor 'Emotion Management' (*Adj.* $R^2 = 0.315$, $p < 0.05$). The team coordination measures based on window 4 were able to significantly predict 25.8% of the variability in the collaboration factor 'Grounding' (*Adj.* $R^2 = 0.258$, $p < 0.05$), 21.0% of the variability in the collaboration factor 'Emotion Management' (*Adj.* $R^2 = 0.210$, $p < 0.05$) and 28.3% of the variability in the collaboration factor 'Emotion Modeling' (*Adj* $R^2 = 0.283$, $p < 0.05$).

Window size 10%

For the window size of 10%, every collaboration factor could be predicted by at least one of the windows to some extent, having an adjusted R^2 value higher than 0. However, for the collaboration factors 'Co-Construction', 'Emotion Modeling' and 'Transactivity' no significant models were generated. Collaboration factors 'Grounding', 'Conflict' and 'Confrontation' each had one window for which the regression model based on that data was able to significantly predict the collaboration factor. The collaboration factor 'Convergence' was significantly predicted by two different regression models based on different windows, and the collaboration factor 'Emotion Management' was significantly predicted by as much as five different regression models based on different windows.

The regression model that significantly predicted 'Grounding' was based on window 8 and accounted for 46.6% of the variability in the data (*Adj.* $R^2 = 0.466$, $p < 0.01$). The regression model that significantly predicted 'Conflict' was based on window 7 and accounted for 33.1% of the variability in the data (*Adj.* $R^2 = 0.331$, $p < 0.01$). The regression model that significantly predicted 'Confrontation' was based on window 7 and accounted for 25.4% of the variability in the data (*Adj.* $R^2 = 0.254$, $p < 0.05$).

Two regression models were able to significantly predict 'Convergence'. The regression model based on window 5 accounted for 24.1% of the variability in the data (*Adj.* $R^2 = 0.254$, $p < 0.05$) , while the regression model based on window 1 accounted for 19.7% of the variability in the data (*Adj.* $R^2 = 0.197$, $p < 0.05$).

Five regression models were able to significantly predict 'Emotion Management'. These were the models based on window 7 (*Adj.* $R^2 = 0.463$, $p < 0.01$), window 2 (*Adj.* $R^2 = 0.321$, $p < 0.01$), window 5 (*Adj.* $R^2 = 0.271$, $p < 0.01$), window 4 (*Adj.* $R^2 = 0.228$, $p < 0.05$) and window 3 (*Adj.* $R^2 = 0.207$, $p < 0.05$).

5. Discussion

One of the goals of this study was to investigate how multimodal team coordination measures compare to single-modality team coordination measures in how well they can predict team performance (RQ1). By comparing the standardized beta coefficients within regression models containing both multimodal and single-modality team coordination measures, it was found that for some collaboration factors multimodal team coordination was more important than single-modality team coordination measures, while for other collaboration factors the opposite was true.

Particularly for the collaboration factors 'Co-Construction', 'Emotion Modeling' and 'Transactivity', multimodal team coordination measures had more predictive power than single-modality team coordination measures. This indicates that the underlying mechanisms that influence these collaboration factors likely take place on the multimodal level more than on single-modality levels. Subsequently, this means that studies interested in these collaboration factors should include both heart rate variability and electrodermal activity measurements and should employ analyses that are suitable for capturing multimodal interaction patterns.

For the collaboration factors 'Grounding', 'Conflict' and 'Convergence', multimodal team coordination had less predictive power than single-modality team coordination measures. This indicates that for these collaboration factors, underlying mechanisms likely take place within a single modality, or at least not across the two modalities that were investigated in this study. Future studies that focus on predicting these collaboration factors do not need to employ analyses that are suitable for capturing multimodal interaction patterns, or they should investigate other modalities that might have an effect on these collaboration factors.

Finally, for the collaboration factors 'Emotion Management' and 'Confrontation', the results were mixed. For both factors, one of the multimodal team coordination measures had a larger effect on the collaboration factor than its single-modality counterparts, while the other multimodal team coordination measure had a smaller effect on the collaboration factor than its single-modality counterparts. A possible explanation for this is that determinism and recurrence measure different aspects of team coordination, and the underlying mechanisms influencing these collaboration factors might be represented multimodally for one of these aspects, while they are not represented on the multimodal level for the other aspect.

The second goal of this study was to investigate how segments of physiological data could be used to predict subjective team performance (RQ2). In order to do so, the effect of segment size of physiological data used for predicting subjective team performance was investigated. Additionally, the effect of the location of the segment in the physiological data was also examined.

It was found that team coordination measures calculated from smaller windows of physiological data were generally able to predict more variability in the subjective team performance measures than team coordination measures calculated from larger windows of physiological data. Additionally, for most of the collaboration factors, the best performing regression models among the models based on 10% windows were able to statistically significantly predict the collaboration factors, while for the 50% windows only 'Emotion Management' could be predicted significantly.

The one exception to this is the 'Emotion Modeling' factor. This factor was predicted significantly using the fourth 20% segment, while no 10% or 50% segments were able to significantly predict this collaboration factor. A possible explanation for why this collaboration factor cannot be predicted significantly by the 10% segments while it can be predicted by a 20% segment, is that recurrence patterns related to 'Emotion Modeling' require more context than the 10% segments are able to provide. If recurrence patterns happening in the 60%-80% range of the data are critical for predicting 'Emotion Modeling', it makes sense that only the 20% window was able to predict this factor.

Since some of the segments appear to be more critical for predicting certain collaboration factors than others, future research into these collaboration factors can focus more on these segments to get better insights into how physiological recurrence is related to these factors. Alternatively, research on subjective team performance can incorporate this knowledge in their design, so that they can make more accurate predictions of subjective team performance. Furthermore, they would only need a limited amount of data to be able to predict subjective team performance. This could make research using physiological data more robust to corrupt data or movement artifacts, because as long as the critical segments are intact, the subjective team performance measures can still be predicted.

6. Conclusion

One of the goals of this study was to investigate how well multimodal team coordination measures could predict subjective team performance in comparison to single-modality team coordination measures. The results of this study showed that for three out of eight collaboration factors multimodal team coordination measures were more important for prediction than the single-modality variants, and for two out of the eight collaboration factors one of the multimodal team coordination measures was better at predicting the collaboration factors than the single-modality variant. Future studies investigating ‘Co-Construction’, ‘Emotion Modeling’, ‘Transactivity’, ‘Emotion Management’ and ‘Confrontation’ should consider adding multimodal coordination measures, as for these factors it appears that underlying mechanisms are embedded at the multimodal level.

The second goal of this study was to investigate the effectiveness of segmented physiological data in predicting subjective team performance. It was found that some critical smaller segments of the physiological data were sufficient for significantly predicting five out of eight collaboration factors. This implies that underlying mechanisms within the investigated modalities that are important for predicting subjective team performance can be located using this approach.

Future research should investigate whether similar results can be found in experiments where teams perform different types of collaborative tasks. Additionally, future research should investigate whether the addition of more modalities could further enhance the ability to predict subjective team performance. For example, the addition of eye-gaze coordination or movement coordination could further capture multimodal patterns related to subjective team performance.

Finally, this study did not look into overlapping segments of physiological data, and only evaluated a few different window sizes. It is possible that the performance of regression models based on team coordination calculated from segments of physiological data would be completely different if different segments were considered, with either larger or smaller sizes, or shifted along the temporal axis by some percentage. A more complete understanding of underlying recurrence patterns affecting the different subjective team performance measures could arise by investigating more variations of data segments. Unfortunately, a more nuanced approach that could look into these differences was outside of the scope of this study.

Self-reflection

This section is a required component of the CSAI Bachelor Thesis.

Scientific Research Process

During the bachelor thesis project, I have learned a lot about the scientific research project. I was initially planning to collect my own data as part of a research project under Travis J. Wiltshire. However, due to the Corona regulations, this experiment was canceled. As a result, I had to come up with a new research proposal, where I would use an existing dataset instead of collecting my own data. While I struggled in finding a suitable dataset for my research questions, my supervisor was able to help me out by suggesting multiple datasets. Unfortunately, my old research questions could not be answered using the available datasets, so I had to change my research questions and plan of analysis in a short amount of time.

Partly as a result of these hasty revisions, I have run into several issues later on in the scientific research process. My initial plan of analysis was not suitable for answering my research questions perfectly, but thanks to the help of my supervisor I was able to find a better way for answering my research questions. Likewise, I found out that some of the theories that I was initially planning to use did not really apply to my new research questions.

In future research projects, I will need to be more careful in formulating my research questions and plan of analysis to avoid running into these kinds of issues. Especially when conducting my own experiment, I should prepare a slightly broader theoretical background, so that I can more easily adjust my research questions. Additionally, I should look into existing datasets more as a backup plan, even when I plan to collect my own data.

Conducting Analyses

I ran into a few issues when conducting my analyses. When I created my research proposal, I did not yet fully understand how the MdRQA technique worked. I only found out later that the different time-series that were used in the MdRQA needed to be of the same length, with the same sampling rate. As a result, I had to adjust my initial plan for calculating heart rate variability and use inter-beat intervals instead of using SDNN measures. Fortunately, this had little impact on my final thesis project, as inter-beat intervals are still a valid measurement for heart rate variability.

In future research projects, I will need to make sure that I have a more thorough understanding of the methods of analysis that I plan to use prior to creating a research proposal. It would especially be a good idea to make sure that I have part of the dataset available before writing the research proposal, so that I can test out the methods of analysis with a small sample of the data.

Poster Presentation

One part of the thesis project was the poster presentation, where we were had to create a poster for our thesis project and present the posters in an exhibition-like setting. Due to the Corona regulations, the exhibition-like setting was canceled, but video presentations were uploaded online instead. I found that the creation of a scientific poster was incredibly difficult and struggled with deciding which information was important enough to include on the poster and which information was redundant. Based on the feedback from my supervisor, as well as by comparing my poster to the posters of classmates, I realized that I had included way too much text on my poster, and my poster looked incredibly bland.

In the future, I will be very selective in which information to add to my research proposal. Rather than include a lot of text, I will try to visualize important aspects of the research. If for future projects I can get access to the dataset earlier on, I would include some better visualizations of the data that I work with.

Writing the Thesis

As a result of the Corona regulations, all students had to work from home for the majority of this semester. I struggled with focusing on my thesis in my home-working environment, and as a result fell behind on my planning for writing the academic report.

In the future I should make sure that my working environment is suitable for writing a thesis. Additionally, I should put more effort into staying on track with my planning, so that I am better able to receive feedback on my work. Finally, if all else fails, I should contact my supervisor sooner so that they can help me out, rather than try to fix everything myself.

7. References

Ahonen, L., Cowley, B., 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.

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.

Bouchequet, P. (2020). rsleep: Analysis of Sleep Data. R package version 1.0.2. Retrieved from. https://CRAN.R-project.org/package=rsleep

Carrère, S., & Gottman, J. M. (1999). Predicting divorce among newlyweds from the first three minutes of a marital conflict discussion. *Family process*, 38(3), 293-301.

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

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.

Eloy, L., EB Stewart, A., Jean Amon, M., Reinhardt, C., Michaels, A., Sun, C., ... & D'Mello, S. (2019, October). Modeling Team-level Multimodal Dynamics during Multiparty Collaboration. *In 2019 International Conference on Multimodal Interaction* (pp. 244-258).

Fusaroli, R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive science*, 40(1), 145-171.

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

Henning, R. A., Armstead, A. G., & Ferris, J. K. (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.

Henning, R. A., & Korbelak, K. T. (2005). Social-psychophysiological compliance as a predictor of future team performance. *Psychologia*, 48(2), 84-92.

Kazi, S., Khaleghzadegan, S., Dinh, J. V., Shelhamer, M. J., Sapirstein, A., Goeddel, L. A., Chime, N. O., Salas, E. & Rosen, M. A. (2019). Team Physiological Dynamics: A Critical Review. *Human factors*, 0018720819874160.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from. https://www.R-project.org/.

Richardson, M., Garcia, R. L., Frank, T. D., Gregor, M., & Marsh, K. L. (2012). Measuring group synchrony: a cluster-phase method for analyzing multivariate movement time-series. *Frontiers in physiology*, 3, 405.

Schoenherr, D., Paulick, J., Strauss, B. M., Deisenhofer, A. K., Schwartz, B., Rubel, J. A., ... & Altmann, U. (2019). Nonverbal synchrony predicts premature termination of psychotherapy for social anxiety disorder. *Psychotherapy*.

Tschacher, W., & Meier, D. (2019). Physiological synchrony in psychotherapy sessions. *Psychotherapy research*, 1-16.

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.

Wallot, S., Roepstorff, A., & Mønster, D. (2016). Multidimensional Recurrence Quantification Analysis (MdRQA) for the analysis of multidimensional time-series: A software implementation in MATLAB and its application to group-level data in joint action. *Frontiers in psychology*, 7, 1835.

Wallot, S., & Leonardi, G. (2018). Analyzing multivariate dynamics using cross-recurrence quantification analysis (crqa), diagonal-cross-recurrence profiles (dcrp), and multidimensional recurrence quantification analysis (mdrqa)–a tutorial in r. *Frontiers in psychology*, 9, 2232.

Appendix A: Standardized Beta Coefficients of the single-modality and multimodal team coordination measures in eight different regression models

Team coordination measurement	Grounding	Conflict	Convergence	Confrontation	Co-construction	Emotion Management	Emotion Modeling	Transactivity
%Recurrence IBI + EDA	0.305	-0.074	0.395	-0.383	-0.262	-0.368	-0.131	-0.401
%Determinism IBI + EDA	0.112	-0.412	0.213	-0.457	-0.504	-0.509	-0.150	-0.452
%Recurrence IBI	0.133	0.080	0.157	0.333	0.631	0.537	0.764	0.917
%Determinism IBI	-0.556	0.159	-0.159	0.313	1.091	0.390	0.369	1.137
%Recurrence EDA	-0.877	0.353	-0.364	0.613	1.619	0.799	0.795	1.385
%Determinism EDA	0.794	-0.285	0.074	-1.001	-2.312	-0.755	-1.077	-2.246