Inter-brain synchrony during (dis)agreement in monastic debate.

Bachelor's Project Thesis

Nathan van Beelen, n.van.beelen@student.rug.nl,
Supervisors: Dr. M. K. van Vugt & Ms. P. Kaushik

Abstract: Research has suggested higher inter-brain synchrony is associated with improved cooperation. A similar effect was later found to also exist for periods of agreement and disagreement in a monastic debate, where agreement led to a higher inter-brain synchrony. As periods of agreement and disagreement can be subjective, this study aimed to confirm this effect of higher inter-brain synchrony during agreement in a more controlled manner. This was done using debate specific vocabulary in a monastic debate as indication for agreement and disagreeement. In order to confirm this hypothesis, the inter-brain synchrony between pairs of monks was measured with the use of EEG hyperscanning. While we did not observe average differences between utterances reflecting agreement and disagreement, the Support-Vector Machines classifier was able to distinguish between the two conditions on a single-trial basis. A feature importance analysis was done on both the SVM and LDA (Linear Discriminant Analysis) classifiers while mitigating the effect of correlation between the different features. No significant differences were found between the importance of different features using the one-way ANOVA. This led to the conclusion that there is indeed a difference of averages in inter-brain synchrony between agreement and disagreement where agreement leads to a decrease in inter-brain synchrony.

1 Introduction

Research done by Cui, Bryant, and Reiss (2012) on cooperation and competition between two participants has found that the inter-brain synchrony between two participants increased during cooperation, but not during competition. This is part of a recent move within the social neuroscience from within-individuals to between-individuals to study a fundamental part of social interactions (Schilbach et al., 2013). Inter-brain synchrony is one such way of measuring the neural activity in social interactions.

Inter-brain synchronisation can be measured using hyperscanning which assesses the level of coupling between two brains (Schoot, Hagoort, & Segaert, 2016). There are several methods of recording neural activity in the brain to perform hyperscanning, such as EEG, (f)NIRS, and fMRI (Sänger, Müller, & Lindenberger, 2012; Miller et al., 2019; Montague et al., 2002). Although EEG appears to be the more popular choice, due to its low cost and wider availability.

Inter-brain synchronisation has been studied in different contexts. Examples of this are group interactions in the classroom (Dikker et al., 2017) and joint actions such as the leader-follower task (Sänger et al., 2012) and the movement synchronisation task (Tognoli, Lagarde, DeGuzman, & Kelso, 2007)

One of the reasons for using inter-brain synchrony as a way to measure the neural activity in social interactions is because inter-brain synchrony might play a role in positive social interactions. An earlier named example of this is cooperation, which was found to lead to an increase in inter-brain synchrony (Cui et al., 2012). No similar effect was found for competition. This is likely caused by the participant having to model and predict the behaviour of others in order to cooperate effectively. Furthermore, synchrony is said to be a critical component of human attachments as it evolved from coordinated group activity (Feldman, 2017).

Research by Mu, Guo, and Han (2016) has shown

Figure 1.1: Monastic debate as practiced by Tibetan monks at Sera Jey monastery (M. K. van Vugt et al., 2020)

that cooperation in a lab setting leads to more inter-brain synchronisation in the alpha band (8-12 Hz). A study done by M. K. van Vugt et al. (2020) tested the difference of inter-brain synchrony between agreement and disagreement in a more naturalistic environment.

The environment used in M. K. van Vugt et al. (2020) is a form of practice in Buddhism called monastic debate (see figure 1.1). Monastic debate is a highly social and interactive form of meditation - as opposed to mindfulness meditation - in order to deepen the participating monks' understanding of the philosophical material studied. In a monastic debate of two participants - but do note that more people can participate - there is the role of challenger and defender. The role of the challenger is to pose questions and statements in such a way that the defender will agree to statements that are contradictory. The role of the defender is to answer the challenger without contradicting themselves. During the debate the defender is limited to four different responses: (1) I agree, (2) please state a reason why, (3) the reason is not established, or (4) no pervasion. The debate ends once the defender contradicts themselves, the challenger is unable to force the defender into contradicting themselves, the challenger breaks down, or a time limit has been reached (Dreyfus, 2008; M. K. van Vugt et al., 2019).

A monastic debate starts with the challenger proposing the topic of the debate. Although the challenger may directly proceed to debate the defender, it is common to first ask explanations of the defender in order to understand his position on the topic. At this point the challenger will begin to set forth consequences to draw the defender to contradict themselves. The defender on the other hand tries to block these contradictions which they can do with the responses given above (M. K. van Vugt et al. (2020), appendix 1).

M. K. van Vugt et al. (2020) was able to find a difference in periods of agreement and disagreement in a monastic debate, where periods of agreement and disagreement were defined by monk ratings. This raises the question whether the difference in inter-brain synchrony is reproducible in a more controlled manner. Namely, when looking at simple debate answers. Therefore, this study looks at the difference in inter-brain synchronisation between the defender answering 'I agree' or 'please state a reason why' (which implies 'no I do not agree'). This leads to the research question: "How is the debate-specific vocabulary used by the debating monks correlated with inter-brain synchrony?"

It is expected that the debate-specific vocabulary used by the debating monks is indeed associated with inter-brain synchrony as this was also found by M. K. van Vugt et al. (2020) when using agreement and disagreement as rated by monks. This way, the following hypothesis was created: "Inter-brain synchrony is positively associated with debate-specific vocabulary used by debating monks." Where positively associated means that there is more inter-brain synchrony when the debating monks are in agreement than when they are in disagreement.

In order to test the research question, a new, yet unpublished, dataset collected by M. K. van Vugt et al. (2020) is used. The dataset was created by measuring the brain activity of two monks participating in a monastic debate by the use of EEG. This EEG can be used for hyperscanning as it was recorded simultaneously (see Cui et al. (2012) for an example of NIRS hyperscanning). Additionally, videos of the debate were recorded. This can be used to extract moments of agreement and disagreement.

Earlier research has already laid out a foundation of the cognitive functions that can be measured by different bands of wavelengths in EEG. The bands and their association are the following: 1) the theta (4-9Hz) oscillations associate with attention, ab-

sorption and cognitive control (Cavanagh, Frank, Klein, & Allen, 2010), more specifically, theta oscillations in predominantly parieto-temporal locations are associated with both accumulating and comparing information (M. K. van Vugt, Simen, Nystrom, Holmes, & Cohen, 2012), and memory encoding and retrieval (Sederberg, Kahana, Howard, Donner, & Madsen, 2003), 2) the alpha (10-14Hz) oscillations associate with idling and inhibition (Händel, Haarmeier, & Jensen, 2011; Pfurtscheller, Stancák, & Neuper, 1996), 3) the beta oscillations (14-28Hz) are mostly associated with motor activity (Brovelli, Ding, Chen, Nakamura, & Bressler, 2004), and 4) faster gamma (28-48Hz) oscillations have been associated with focused attention (Bauer, Oostenveld, Peeters, & Fries, 2006; Hoogenboom, Schoffelen, Oostenveld, Parkes, & Fries, 2006). Research on cooperation and competition, which is similar to the focus of this study, has found that alpha and theta bands play an important role on the centro-parietal and centro-frontal regions respectively (Balconi & Vanutelli, 2016; Hu et al., 2018). Furthermore, research by M. K. van Vugt et al. (2020) also found the alpha band to indicate significant differences between agreement and disagreement in the frontal region. In all instances, cooperation or agreement led to the increase of inter-brain synchrony. In light of these findings I will also look at the oscillatory power within-individuals on the theta and alpha band to examine if these effects can also be found in the current study.

Being able to classify states of agreement and disagreement can prove useful both in future studies as well as more practical applications. In future studies classification can help answer questions such as whether the debate-specific utterances of the participating monks or the ratings of observing monks are more reliable indicators of agreement and disagreement. On a more practical level, classification helps in classifying data that has to be investigated (e.g. videos don't have to be manually tagged anymore) and the possibility to do online classification may arise.

Although the use of classification in EEG has been studied before (Vézard, Legrand, Chavent, Faïta-Aïnseba, & Trujillo, 2015; Zhao et al., 2018; Jrad & Congedo, 2012), the use of classification on the topic of hyperscanning or cooperation was not very prevalent (Verdiere, Dehais, & Roy, 2019). Classification methods that were used include a variant of Linear Discriminant Analysis (LDA) (Tharwat, Gaber, Ibrahim, & Hassanien, 2017): shrinkage LDA (Ahdesmaki & Strimmer, 2012), Support-Vector Machines (SVM) (Cortex & Vapnik, 1995) and extreme learning machine (ELM) (Huang, Zhu, & Siew, 2007). The differentiation between agreement and disagreement is a binary classification task. As such, all three methods are suitable for binary classification. All three methods are used to train on the EEG data from the monastic debates and the labels assigned to these data. After which the classifiers were used to predict agreement and disagreement on a single-trial basis.

2 Methods

In order to answer how debate-specific vocabulary correlates with inter-brain synchrony, EEG recordings of the monks participating in a monastic debate have been taken. With the use of these recordings, I have studied the difference in inter-brain synchrony during agreement and disagreement. I used these findings as features to train classifiers in order to find future occurrences of agreement and disagreement.

The data that I use in this study was collected in the study by M. K. van Vugt et al. (2020). For clarity, I summarise the methods of the experiment here.

2.1 Participants

The participants were all Tibetan monks from the Sera Jey monastery aged between 20 and 30 years old. All participants were male. The participants were recruited by way of an announcement in the monastery. Additionally, a few monks were asked directly. In order to limit the difference between experienced and inexperienced monks M. K. van Vugt et al. (2020) endeavored to select students with top marks in their classes. The participants were rewarded by being served lunch or dinner following their participation. There were two groups of monks. One group existed out of experienced monks, which were monks on the Vinaya class level (equaling at least fifteen years ($\sim$18750 hours) or more of experience). The other group existed out of

Figure 2.1: Monastic debate as they were being recorded with the use of EEG measurement devices.

inexperienced monks which were from the Paramita class level (with at least three years (∼3750 hours) of experience). The experienced monks numbered a total of 10 while the inexperienced group had 14 monks.

2.2 Design

The monastic debate took place between a pair of monks. One of the monks played the role of challenger while the other played the role of defender. An audience was present during the debates (see figure 2.1) but they were not allowed to interfere. Two different debates were held: easy and hard, where the specific topic of debate was manipulated in difficulty.

In this study I follow a within-subject design where every utterance of agreement and disagreement represents a trial.

2.3 Procedure

The experiment is started with the easy debate and followed up by the hard debate.

Before the experiment was started, the participants were told about the procedure. They were told that participation in the study was completely voluntary and that they had the possibility to quit at any point of time without repercussions. At which point the participants gave verbal informed consent. The study was conducted in accordance with the declaration of Helsinki.

The topic of the debates was "The Definition of Bodhicitta". The reason that this topic was chosen is because of its familiarity among both inexperienced and experienced monks. Since the experienced monks had studied this topic many years ago, all monks were instructed to review their textbook for 15 minutes before the debate started.

Before the actual start of the debate the participants provided personal information such as their age, the year they started their studies at the monastery, and their level in monastic training. This information was tied to a sequential number serving as their identification number in order to anonymise the data. At this point the EEG caps were applied and the easy debate was performed followed by the hard debate (see figure 2.1 for an impression of what the setup looked like). All participants played at least one role in the debate. If time allowed for it, the roles were reversed and the experiment was repeated. This was the case for almost all instances.

2.4 Video recordings

In order to know at which points during the EEG recordings the words 'I accept' and 'Why?' were said, video recordings were made at the same time as the EEG recordings. In total 44 videos were recorded which all contained one debate (either easy or hard) per video. The easy debates took around 10 minutes while the hard debates took around 15 minutes, creating an average of about thirteen minutes per video. In all videos, there was only one challenger and one defender. 17 videos included experienced monks while 27 videos included inexperienced monks. The videos were recorded with a video camera that had audio recording capabilities. The sampling frequency of the audio was 48KHz. The synchronisation of the video recordings with the EEG measurements were initially done by indicating the start of the EEG measurements in the video. In the last 14 videos a different method was used. One participant would blink their eyes five times in front of the video camera. This would result in five easily recognisable blinking artifacts in the EEG data which would allow us to compute the temporal offset between the video and the EEG streams. This was done in order to minimise the discrepancy in synchronisation between the video and EEG streams.

Annotation	Easy	Hard	Total
'I agree'	3.83	8.27	270
'Why?'	0.35	2.14	55

Table 2.1: An overview of the average number of annotations established as ground truth per video grouped on debate category. The last column notes the total number of annotations made for that category.

2.4.1 Video analysis

The analysis of the videos was done by three researchers using the BORIS event-logging software (Friard & Gamba, 2016). Each researcher would watch the video independently and tag the times at which the keywords indicating agreement or disagreement were uttered. Any tags that were in common would be established as ground truth. In order for a tag to be in common, a tag of one researcher would have to be within a one second range of the tag of another researcher, and the tag would have to be of the same category.

A delay of roughly half a second was introduced in the annotations as human response time is not instanteneous. For this reason all trials were moved back by half a second before the analysis.

A categorisation of the annotations can be seen in table 2.1. A disparity between the number of annotations of hard and easy videos can be noted. This reflects the difference in difficulty between both types of debates.

2.5 EEG recordings

The EEG measurements of the participants were taken using two Biosemi EEG headsets (one for each participant participating in the debate) with 32 electrodes arranged in the international 10-20 system. Both monks participating in the monastic debate were recorded simultaneously. The sampling rate of the electrodes in the EEG headset is 256Hz but was recorded through a bandpass-filter of 0.1-1000Hz. The electrodes were adjusted until the electrical impedances were below 25 kΩ.

2.6 Preprocessing

Movement is an essential part of the monastic debate and makes it more engaging. An example of this is stomping on the ground after the challenger makes a statement. Due to this and other movements, a lot of EEG artifacts were created. However, EEG artifacts were not just created by (explicit) movement. Other causes of EEG artifacts include inexplicit movements such as eye blinks, and jaw-clenching, but also heart beats. Artifacts were removed as they are not a byproduct of brain activity and are not relevant for the analysis. Brain artifacts were identified and removed using independent componenent analysis (ICA) (Comon, 1994). ICA is able to split the signal into different components (much like the ear). EEG artifacts such as heart beats have very different temporal dynamics than EEG itself. This makes it possible to identify these sources of EEG artifacts as independent sources and remove them, before transforming the signal back into the original space. This way the signal that is used for analysis is corrected.

The EEG data was cut in segments with an interval of two seconds. One interval exists out of 512 samples (256 x 2) across 32 channels of EEG data. The power and cross-correlation at each channel was computed by centering the windows around values found at a distance of 100ms. For 2 seconds of EEG - which is the length of each trial - this resulted in 10 samples of cross-correlation and power spectra across 32 channels.

2.7 Feature extraction

In order to perform time-frequency analysis, the Fieldtrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) in Matlab (MATLAB, 2019) was used. Raw EEG data was converted by applying `ft_frequencyanalysis()` for a given frequency interval. This resulted in the power spectrum for all channels of that given frequency. The power spectra was then calculated using the `mtmconvol()` function from the Fieldtrip toolbox. In order to prevent spectral leakage, DPSS (discrete prolate spheroidal sequences) were used as tapers. Multiple tapers were used as they have the advantage of being able to detect non-stationary signals (M. van Vugt, Sederberg, & Kahana, 2007). Which is useful considering the way signals are generated in the

brain (Kaplan, Fingelkurts, Fingelkurts, Borisov, & Darkhovsky, 2005).

2.8 Statistics

The data analysis was carried out using the Matlab programme (MATLAB, 2019) with the use of the Fieldtrip toolbox (Oostenveld et al., 2011). The EEG data were first frequency-transformed. This was done by convolving it with a Hanning taper as the bands of interest were below 30Hz. I defined the theta band to contain the frequencies between 4 and 8 Hz (Sederberg et al., 2003), and the alpha frequency to contain the frequencies between 9 and 13Hz (Händel et al., 2011).

For all trials the inter-brain synchrony of both participants and the oscillatory power for each individual participant was computed in a within-trials design for every channel and for the frequency bands of interest (M. K. van Vugt et al., 2020). This was then compared using linear mixed effects models (Pinheiro & Bates, 2004). I made use of the linear mixed effects models as it is robust to violations of independence (Baayen, Davidson, & Bates, 2008). Furthermore, it is less susceptible to individual differences between participants and the statistical power is higher than that of a regular ANOVA. The baseline used for the t-statistics in the linear mixed effects models is 0.

As a different model was made for each combination of channel and frequency band, the chance of making a type-I error increases. In order to counteract this, the False Discovery Rate (FDR) method (Benjamini & Hochberg, 1995) was used to calculate the correction in p-value. The False Discovery Rate method was used as opposed to the (Holm-)Bonferroni method (Holm, 1979) as the risk of making a type-I error could be manually controlled. The cut-off Q-value of the FDR was set to 0.05, limiting the chance of significant values that are false positives to 5% in multiple-measurements testing.

Outliers were defined as being more than 4 standard deviations removed from the mean. All outliers were removed from the dataset.

2.9 Classification

Classifiers were used to detect if I could discriminate between utterances of agreement and disagreement. The classifiers were trained on all channels, but different instances of the classifiers were used for different data sources. These data sources were a combination of inter-brain synchrony, oscillatory power of the challenger, or oscillatory of the defender and the theta or alpha frequency band.

In the classification selection, a few characterisations were preferable. First of all, a supervised classifier was needed as I wanted the classifier to classify the data based on the two labels (agreement and disagreement). This also means that I am working with a binary output. Furthermore, our data had a low number of samples and a high number of dimensions.

A possible classifier that is well-grounded in both the neuroscience and machine learning literature (Mechelli & Vieira, 2019; Wei et al., 2018; Bayram, Kizrak, & Bolat, 2014; Saccà, Campolo, Mirarchi, & Gambardella, 2018) is the Support-Vector Machines (SVM) classifier (Cortex & Vapnik, 1995). SVM is able to work well with a high-dimensionality of data, this is also true when working with a low number of samples. Another possibility is the shrinkage Linear Discriminant Analysis (shrinkage LDA) (Tharwat et al., 2017; Ahdesmaki & Strimmer, 2012). Like SVM it is able to work well with a high number of dimensions, even when using a low number of samples, due to the shrinkage adaptation. Additionally, unlike SVM it has feature selectivity. Finally, a third classifier that has been used to classify EEG data in the literature is the Extreme Learning Machine (ELM) (Huang et al., 2007). Due to the random initialisation of the hidden layer, it is able to get a reasonably good performance with a short training time. This is an advantage over LDA, which is not able to deal well with a low number of samples (although this is somewhat mitigated by the shrinkage adaptation). All three classifiers will be tested in this study.

The classifiers were implemented in the Python programming language (van Rossum, 1995), using the scikit-learning (Pedregosa et al., 2011) machine learning library. As the ELM algorithm wasn't implemented in scikit-learning, a separate library was used (Lambert, 2013). I implemented the SVM classifier by using cross-validation on the training data for estimating the C and gamma parameters. I used the Gaussian radial basis function as kernel as it is very generalisable (no prior knowledge about the data was needed). In the implementation of LDA I used the least squares solution solver as it can be

combined with shrinkage and it does not need to calculate the covariance matrix (which would make it unsuitable for usage with a high number of features). The shrinkage itself is determined by using the Ledoit-Wolf lemma. For the implementation of ELM I used the traversal method to find a good number of hidden neurons. The activation function used is the hyperbolic tangent function:

$$tanh(x) = \frac{sinh(x)}{cosh(x)} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

In all implementations k-fold cross-validation was used in order to prevent overfitting.

The classifiers were tested by using k-fold cross-validation with $k = 10$. The performance was measured by the accuracy, precision and recall. As precision and recall are both measurements based on the relevance of the data, the F1 score is used as an additional measurement of the quality of the classifications for better interpretability as it turns precision and recall into one score of quality.

2.9.1 Feature importance analysis

In order to see what features are used by the classifiers to classify the testing data, two different feature importance analyses were performed. The first feature importance analysis will make use of an algorithm that trains the classifier once with all features and n times with $n - 1$ features where each time a different feature will be removed. This way the baseline performance (where all features are used) can be compared to the performance when a feature is not used. This makes it possible to see whether a feature plays an important role in the classifier. In the second feature importance analysis a similar approach is used, but instead of removing the feature, its values were shuffled in the testing phase. This analysis is also known as permutation importance (Breiman, 2001).

One problem that arises when doing feature importance analysis is that it may undercut the importance of a feature when it correlates with other features. In order to mitigate this, the features were organised in hierarchical clusters based on the correlation distance calculated with Spearman rank-order correlation coefficient (Zwillinger & Kokoska, 2000). These clusters were merged until 10 clusters remain. One feature from each cluster was picked and used in the classifier for the feature importance analysis.

In order to determine whether certain features are important for the classifier to train on, a one-way ANOVA (Heiman, 1997) was performed for every feature importance analysis. This establishes whether a difference in change in performance (against the baseline) can be measured among the different features.

3 Results

I examined whether there was a difference in inter-brain synchrony between agreement and disagreement, and what frequency bands show such a difference. In addition, I asked whether there were differences in oscillatory power between agreement and disagreement for both the defender and the challenger.

Furthermore, different classifiers were used in order to test whether it is possible to distinguish between the agreement and disagreement utterances on a single-trial basis.

3.1 Average differences in brain activity between agreement and disagreement

I first examined if there was a difference in inter-brain synchrony between agreement and disagreement using LMEs. A topography plot of the results can be seen in figure 3.1. When considering a false-discovery rate of 0.05, no significant values were found. However, when uncorrected using a p-value threshold of 0.01, the channels Pz ($estimate = -0.547$, $t(320) = -2.745$, $p = 0.007$), and PO4 ($estimate = -0.5098$, $t(320) = -2.682$, $p = 0.008$) on the alpha band show a significant difference. In these channels, inter-brain synchrony is smaller for agreement than for disagreement.

I also examined the difference in within-individual oscillatory power between disagreement and agreement. I fitted LMEs for both the power of the challenger and the defender. A topography plot of the results can again be seen in figure 3.1. When considering a false-discovery rate of 0.05, no significant values were found. However, when uncorrected using a p-value threshold of 0.01, the channels CP1 ($estimate = 0.1727$, $t(318) = 2.696$, $p = 0.007$),

Pz ($estimate = 0.2674$, $t(318) = 3.265$, $p = 0.001$), and CP2 ($estimate = 0.1533$, $(t = 315) = 2.98$, $p = 0.003$) on the theta band show a significant difference for the power of the defender. This could potentially indicate a difference of power for agreement compared to disagreement in the defender.

3.2 Classification

As significant average differences indicated by the LMEs do not equal being able to predict on a single-trial level, I trained and tested three different classifiers on six different datasets.

These datasets are the same datasets that were used for calculating the statistics, namely: inter-brain synchrony (IBS), power in the defender (Pow. Def.), and power in the challenger (Pow. Chal.) on both the theta and alpha band. However, as the data was unbalanced (269 agreement samples vs. 55 disagreement samples), the agreement samples were subsampled and the disagreement samples were supersampled to create a balanced dataset of 110 samples for both conditions. The data was then seperated based on the frequency band resulting in six different pools of data. This makes it possible to see what data source results in better performance.

In the end, a feature importance analysis was performed to test what channels were important for a classifier to perform well on a single-trial level.

3.2.1 Support Vector Machines

The first classifier used is the Support Vector Machines (SVM) classifier. The hyperparameters C and gamma were estimated individually for each data source, using an exhaustive grid search. Both parameters were in the log space: 1^x with the ranges $-2 \leq x \leq 13$ and $-9 \leq x \leq 3$ for C and gamma respectively. The step size of x was 1 for both parameters.

The average accuracy of the SVM was 0.92 when tested on all different datasets. The distribution of accuracy scores can be seen in figure 3.2. A further comparison of the scores among the different data sources can be seen in figure 3.3. Table 3.1 shows the performance details on all data source. Both the figure and table show that there is very little difference in performance between the different data sources. One thing that should be noted is that the recall score is (almost) 1.0, meaning that it almost always classifies the 'Why?' annotation correctly. This is likely due to the decision boundary leaving more space (for error) for the 'Why?' annotations than for the 'Accept' annotations.

Data source	Acc.	Precision	Recall	F1
θ IBS	0.93	0.88	1.0	0.93
θ Pow. Def.	0.93	0.88	1.0	0.94
θ Pow. Chal.	0.92	0.87	1.0	0.93
α IBS	0.92	0.86	1.0	0.93
α Pow. Def.	0.92	0.87	1.0	0.93
α Pow. Chal.	0.91	0.86	0.99	0.92

Table 3.1: Performance of the SVM classifier on testing data for all different data sources.

Data source	Acc.	Precision	Recall	F1
θ IBS	0.63	0.63	0.63	0.63
θ Pow. Def.	0.61	0.64	0.49	0.56
θ Pow. Chal.	0.53	0.53	0.52	0.52
α IBS	0.62	0.63	0.58	0.60
α Pow. Def.	0.65	0.73	0.49	0.59
α Pow. Chal.	0.49	0.48	0.36	0.41

Table 3.2: Performance of the shrinkage LDA classifier on testing data for all data sources. Showing a slightly above odds performance for all data sources with the exception of the power of the challenger on both the theta and alpha band. The classifier performs no better than chance on these data sources and on the alpha band the recall is even below chance.

3.2.2 Shrinkage Linear Discriminant Analysis

The second classifier that was used is the shrinkage LDA. The average accuracy of the shrinkage LDA was 0.59 when tested on all different datasets. The distribution of accuracy scores can be seen in figure 3.2. A difference in performance can be seen between the different data sources. Both figure 3.3 as well as table 3.2 show a performance not better than odds for the power of the challenger in both the theta and alpha band.

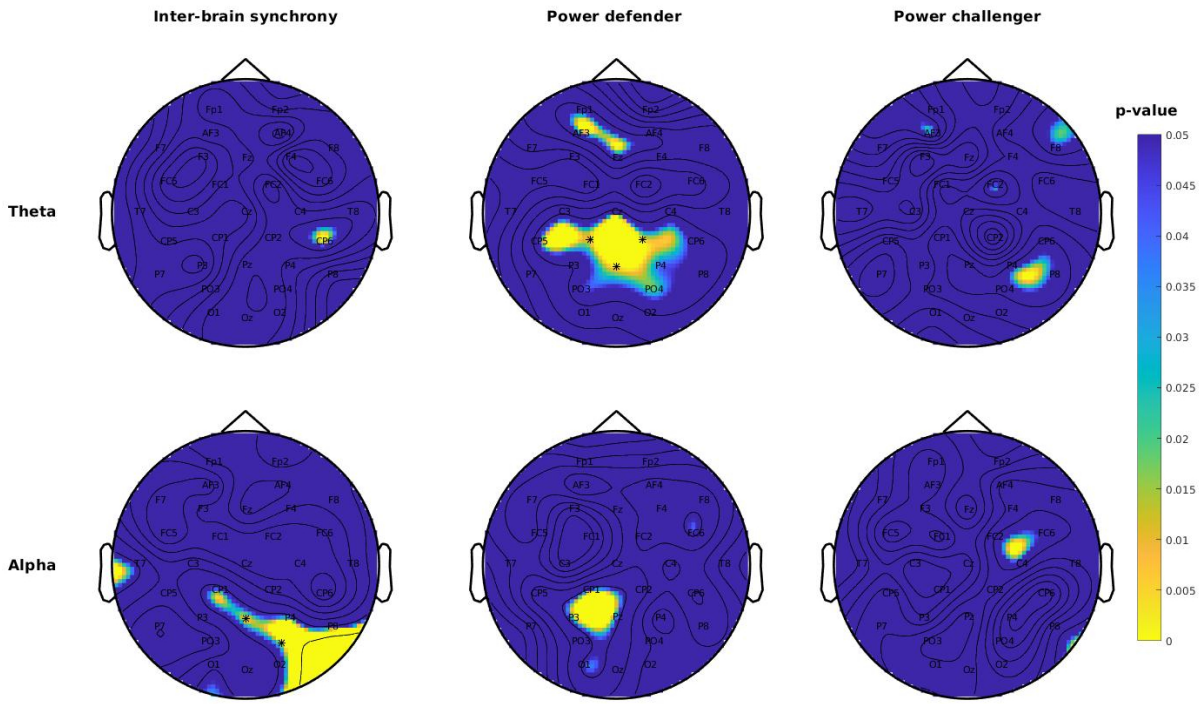

Figure 3.1: Topographical plot of difference between agreement and disagreement in terms of significance for all channels, as measured by the Linear Mixed Effects model. The gradient from blue to yellow indicates smaller p-values, where yellow is significant.

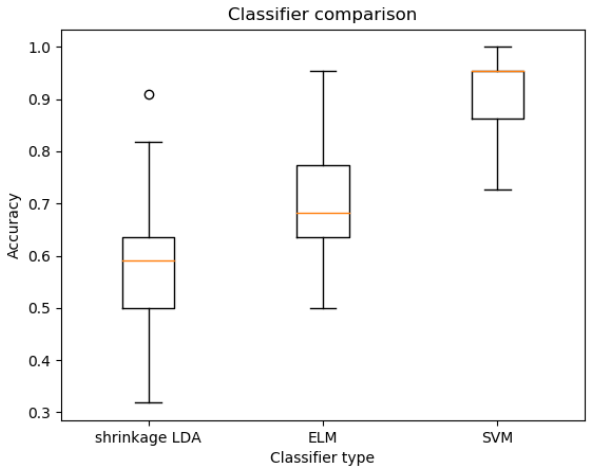

Figure 3.2: Comparison of the performance across all data sources grouped by classifier.

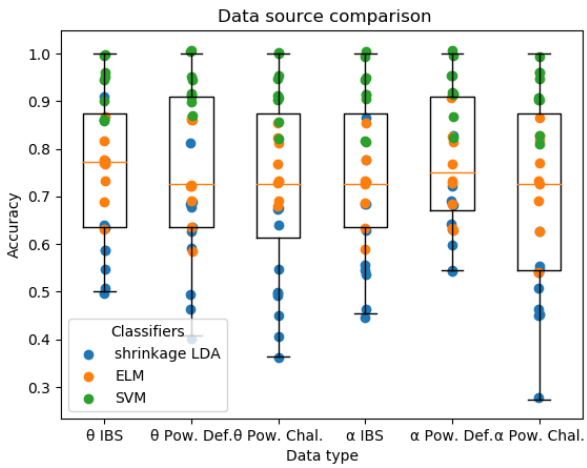

Figure 3.3: Comparison of the performance across the different data sources. The boxplots visualise the performance of all classifiers grouped together. The individual scores of k-fold cross validation is indicated by dots, which are colour-grouped by classifier.

Data source	Acc.	Precision	Recall	F1
θ IBS	0.76	0.9	0.58	0.71
θ Pow. Def.	0.74	0.87	0.58	0.69
θ Pow. Chal.	0.76	0.87	0.61	0.71
α IBS	0.72	0.83	0.55	0.66
α Pow. Def.	0.74	0.82	0.64	0.72
α Pow. Chal.	0.73	0.81	0.58	0.68

Table 3.3: Performance of the ELM classifier on all data sources.

3.2.3 Extreme Learning Machine

The third classifier used is Extreme Learning Machine (ELM). In order to find a good number of hidden neurons, the traversal method was used. The range of possible hidden neurons had a limit equal to the number of samples in the data to minimise overfitting, and a step of 1. The traversal method was used for each data source individually.

The average accuracy of ELM was 0.74 when tested on all different datasets. The distribution of accuracy scores can be seen in figure 3.2. ELM performs above odds, and although it does not have a better performance than SVM, it does perform better than shrinkage LDA as is visualised in figure 3.3. Table 3.3 shows the different performance measurements for all data sources. There does not seem to be a clear difference between different data sources.

3.2.4 Classifier feature importance analysis

Two different feature importance analyses were performed on the shrinkage LDA and SVM classifiers. Shrinkage LDA was used due to its sensitivity to the variance of different channels, however it is offset by its performance which is not much better than odds. SVM on the other hand has a good performance but also a clear decision boundary, making it less ideal for feature importance analysis.

The Spearman rank-order correlation coefficients of the features were calculated for all different sources of data (inter-brain synchrony, power challenger and power defender) on both bands. This was followed up by hierarchical clustering. An example of this can be seen in figure 3.4, where the hierarchy is shown for inter-brain synchrony on the theta band.

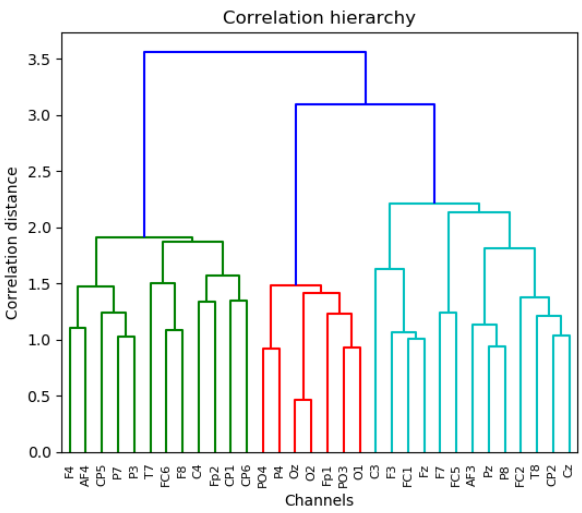

Figure 3.4: A dendogram of the correlations between the inter-brain synchrony of different channels on the theta band in a hierarchical manner. The lower two (or more) channels merge, the more correlated these channels are.

The hierarchical clustering was used in order to group the features with the highest correlations in the same cluster. Clusters were merged until 10 clusters remained. This leads to the minimum correlation distances of 1.55, 0.7 and 1.1 for inter-brain synchrony, power defender, and power challenger on the theta band respectively, and 1.7, 0.8, and 1.1 for inter-brain synchrony, power defender, and power challenger on the alpha band respectively. One feature was picked from every cluster and the feature importance analysis was done on these remaining ten channels.

In the first feature importance analysis, the change in performance was tested by leaving one feature out and testing the performance against the baseline (of all features). In the second feature importance analysis, permutation importance (Breiman, 2001) was used. For all combinations of data sources, frequency bands, and classifiers a one-way ANOVA (Heiman, 1997) was performed on the feature importance analysis to see if there was a difference in importance among the channels. These results can be found in table A.1 and A.2 for the first and second analysis respectively. No signficant differences were found. The p-values were not corrected for multiple measurements.

4 Discussion

This study extends previous inter-brain synchrony research about agreement and disagreement by looking at inter-brain synchrony during utterances of agreement and disagreement in a monastic debate. To reach this goal a number of analyses were performed.

In the first analysis I made use of linear mixed effects models to find a statistical difference between agreement and disagreement and the rate of synchrony. Although numerous trends for a decrease in inter-brain synchrony were found, none of these trends were statistically significant.

This result was unexpected as M. K. van Vugt et al. (2020) found an increase in inter-brain synchrony during periods of agreement compared to periods of disagreement. The trends found in this study on the other hand indicate a decrease in inter-brain synchrony for agreement. However, this could potentially be explained by another result found in M. K. van Vugt et al. (2020). As the channels that were found to be sensitive to monastic experience showed a decrease in inter-brain synchrony during periods of agreement when compared to periods of disagreement. These channels (Fp, CP1, and partially Pz) partly overlap with the channels found to indicate a trend of significance in this study.

Additionally I examined whether there was a significant difference between agreement and disagreement in within-individual oscillatory power. Numerous trends towards a significant difference between agreement and disagreement and oscillatory were found, but none of these proved to be statistically significant.

In the second analysis I looked into whether it would be possible to distinguish between the 'Accept' and 'Why?' utterances on a single-trial basis. Of the three different classifiers that were used to reach this goal, the Support-Vector Machines classifier was able to distinguish between agreement and disagreement with an average accuracy of 92%. This was expected as it corresponds both to previous research by M. K. van Vugt et al. (2020) and Hu et al. (2018). Subsequently, different algorithms were employed to find the channels that lead to this accuracy, but none were found. A potential explanation for the inability to clearly pinpoint certain channels both here and in the statistical analysis might be that the difference between agreement

and disagreement is carried by an interaction effect among multiple channels. This would in turn also explain why the SVM classifier could detect a difference while the LMEs could not.

This study has several potential issues. One such issue is the synchronisation between the video footage that was recorded and the EEG measurements. As the annotations were based on the time denoted in the video footage, this could lead to some discrepancy between the annotation time and the EEG measurement time. After several sessions, this discrepancy was reduced by having the participant blink in front of the camera before the experiment. This is very visible both on camera and in the EEG data and could thus be used as synchronisation point.

Another issue was the language gap between the monks that participated in the debate and the researchers that annotated the videos. Although the researchers were informed of the utterances that they should pay attention to, in some cases this was not clearly audible. An analysis of the annotations further confirmed this issue, as only about 10% of all annotations were agreed upon by all researchers. Context would give more clarity about whether the monks agreed or disagreed in circumstances where that would otherwise be more difficult.

In regards to the research question, the hypothesis predicted a positive association in inter-brain synchrony between disagreement and agreement. Based on the results achieved by the usage of machine learning, the hypothesis was indeed correct in predicting that there is an association in inter-brain synchrony between agreement and disagreement. However, whether this association is positive or negative remains disputable, as feature importance analyses of the classifiers did not yield any results and the results of the LMEs seem to hint at a decrease of inter-brain synchrony in the Pz and PO4 channels - although it was not found to be significant.

Future research can be headed in two directions: the further investigation of inter-brain synchrony or further research into the application of machine learning to distinguish between different mental states. On the topic of further investigation of inter-brain synchrony, it would be good to look into possible causes of the confliciting results between the statistics in this study and the results in M. K. van Vugt et al. (2020). Possible causes can include a difference in timing (e.g. the annotations in this study could be timed earlier and therefore mostly miss the desired effect), but also the fact that a more holisitc approach was used in M. K. van Vugt et al. (2020). On the topic of machine learning, the findings in this study could be further expanded upon by attempting to distinguish between the 'neutral' state and the agreement and disagreement states. This would be relevant for further research into the development of a brain-computer interface that would be able to classify the different mental states online.

References

Ahdesmaki, M., & Strimmer, K. (2012). Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. *Annals of Applied Statistics*, *6*, 503-519. doi: doi:10.1214/09-AOAS277

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi: doi:10.1016/j.jml.2007.12.005

Balconi, M., & Vanutelli, M. (2016). Competition in the brain. the contribution of EEG and fNIRS modulation and personality effects in social ranking. *Frontiers in Psychology*, *7*, 1587-1601. doi: doi:10.3389/fpsyg.2016.01587

Bauer, M., Oostenveld, R., Peeters, M., & Fries, P. (2006). Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. *Journal of neuroscience*, *26*, 490-501. doi: doi:10.1523/JNEUROSCI.5228-04.2006

Bayram, K. S., Kizrak, M. A., & Bolat, B. (2014, 06). Classification of EEG signals by using support vector machines.. doi: doi:10.1109/INISTA.2013.6577636

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*, 289-300. doi: doi:10.1111/j.2517-6161.1995.tb02031.x

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32. doi: doi:10.1023/A:1010933404324

Brovelli, A., Ding, M., Chen, Y., Nakamura, R., & Bressler, S. (2004). Beta oscillations in a large-scale sensorimotor cortical network: Directional influences revealed by granger causality. *Proceedings of the National Academy of Sciences*, *101*, 9849-9854. doi: doi:10.1073/pnas.0308538101

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, *49*, 3198-3209. doi: doi:10.1016/j.neuroimage.2009.11.080

Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, *36*, 287-314. doi: doi:10.1016/0165-1684(94)90029-9

Cortex, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273-297. doi: doi:10.1023/A:1022627411411

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage*, *59*, 2430-2437. doi: 10.1016/j.neuroimage.2011.09.003

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintocka, J., ... J. J. Van Bavel, D. P., M. Ding (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, *27*, 1375-1380. doi: doi:10.1016/j.cub.2017.04.002

Dreyfus, G. B. (2008). What is debate for? the rationality of tibetan debates and the role of humor. *Argumentation*, *22*, 43-58. doi: doi:10.1007/s10503-007-9079-2

Feldman, R. (2017). The neurobiology of human attachments. *Trends in cognitive sciences*, *21*, 80-99. doi: doi:10.1017/S0140525X12000660

Friard, O., & Gamba, M. (2016). Boris: a free, versatile open source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*, 1325-1330. doi: doi:10.1111/2041-210X.12584

Händel, B. F., Haarmeier, T., & Jensen, O. (2011). Alpha oscillations correlate with the successful inhibition of unattended stimuli. *Journal of cognitive neuroscience*, *23*, 2494-2502. doi: doi:10.1162/jocn.2010.21557

Heiman, G. W. (1997). Understanding research methods and statistics: An integrated introduction for psychology..

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65-70. doi: doi:10.2307/4615733

Hoogenboom, N., Schoffelen, J. M., Oostenveld, R., Parkes, L. M., & Fries, P. (2006). Localizing human visual gamma-band activity in frequency, time and space. *NeuroImage*, *29*, 764-73. doi: doi:10.1016/j.neuroimage.2005.08.043

Hu, Y., Pan, Y., Shi, X., Cai, Q., Li, X., & Cheng, X. (2018). Inter-brain synchrony and cooperation context in interactive decision making. *Biological Psychology*, *133*, 54-62. doi: doi:10.1016/j.biopsycho.2017.12.005

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2007). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*, 489-501. doi: doi:10.1016/j.neucom.2005.12.126

Jrad, N., & Congedo, M. (2012). Identification of spatial and temporal features of EEG. *Neurocomputing*, *90*, 66-71. doi: doi:10.1016/j.neucom.2012.02.032

Kaplan, A., Fingelkurts, A. A., Fingelkurts, A. A., Borisov, S. V., & Darkhovsky, B. S. (2005). Nonstationary nature of the brain activity as revealed by EEG/MEG: Methodological, practical and conceptual challenges. *Signal Processing*, *85*, 2190-2212. doi: doi:10.1016/j.sigpro.2005.07.010

Lambert, D. C. (2013). *Python-ELM.* `https://github.com/dclambert/Python-ELM`. GitHub.

MATLAB. (2019). *9.7.0.1190202 (r2019b)*. Natick, Massachusetts: The MathWorks Inc.

Mechelli, A., & Vieira, S. (2019). *Machine learning: methods and applications to brain disorders.* Cambridge, Massachusetts: Academic Press. doi: doi:10.1016/C2017-0-03724-2

Miller, J. G., Vrtička, P., Cui, X., Shrestha, S., Hosseini, S. M. H., Baker, J. M., & Reiss, A. L. (2019). Inter-brain synchrony in mother-child dyads during cooperation: An fNIRS hyperscanning study. *Neuropsychologia*, *124*, 117-124. doi:

doi:10.1016/j.neuropsychologia.2018.12.021

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., ... Fisher, R. E. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *NeuroImage*, *16*, 1159-1164. doi: doi:10.1006/nimg.2002.1150

Mu, Y., Guo, C., & Han, S. (2016). Oxytocin enhances inter-brain synchrony during social coordination in male adults. *Social cognitive and affective neuroscience*, *11*, 1882-1893. doi: doi:10.1093/scan/nsw106

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysical data. *Computational Intelligence and Neuroscience*, *2011*. doi: doi:10.1155/2011/156869

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830. doi: doi:10.5555/1953048.2078195

Pfurtscheller, G., Stancák, A., & Neuper, C. (1996). Event-related synchronization (ers) in the alpha band - an electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology*, *24*, 36-46. doi: doi:10.1016/S0167-8760(96)0066-9

Pinheiro, J. C., & Bates, D. M. (2004). *Mixed-effects models in S and S-PLUS*. New York: Springer. doi: doi:10.1007/b98882

Saccà, V., Campolo, M., Mirarchi, D., & Gambardella, A. (2018, 08). On the classification of EEG signal by using an SVM based algorithm. In (p. 271-278). doi: doi:10.1007/978-3-319-56904-$8_2$6

Sänger, J., Müller, V., & Lindenberger, U. (2012). Intra- and interbrain synchronization and network properties when playing guitar in duets. *Frontiers in human neuroscience*, *6*, 312-331. doi: doi:10.3389/fnhum.2012.00312

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schilicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*, 393-462. doi: 10.1017/S0140525X12000660

Schoot, L., Hagoort, P., & Segaert, K. (2016). What can we learn from a two-brain approach to verbal interaction? *Neuroscience and Biobehavioral Reviews*, *68*, 454-459. doi: doi:10.1016/j.neubiorev.2016.06.009

Sederberg, B., Kahana, M. J., Howard, M. W., Donner, E. J., & Madsen, J. R. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience*, *23*, 10809-10814. doi: doi:10.1523/JNEUROSCI.23-34-10809.2003

Tharwat, A., Gaber, T., Ibrahim, A., & Hassanien, A. E. (2017). Linear discriminant analysis: A detailed tutorial. *AI Communications*, *30*, 169-190. doi: doi:10.3233/AIC-170729

Tognoli, E., Lagarde, J., DeGuzman, G., & Kelso, J. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences of the United States of America*, *19*, 8190-8195. doi: doi:10.1073/pnas.0611453104

van Rossum, G. (1995, May). *Python tutorial* (Tech. Rep. No. CS-R9526). Amsterdam: Centrum voor Wiskunde en Informatica (CWI).

van Vugt, M., Sederberg, P., & Kahana, M. (2007). Comparison of spectral analysis methods for characterizing brain oscillations. *Journal of Neuroscience Methods*, *162*, 49-63. doi: https://doi-org.proxy-ub.rug.nl/10.1016/j.jneumeth.2006.12.004

van Vugt, M. K., Moye, A., Pollock, J., Johnson, B., Bonn-Miller, M. O., Gyatso, K., ... Freso, D. M. (2019). Tibetan buddhist monastic debate: psychological and neuroscientific analysis of a reasoning-based analytical meditation practice. *Progress in Brain Research*, *244*, 233-253. doi: doi:10.1016/bs.pbr.2018.10.018

van Vugt, M. K., Pollock, J., Johnson, B., Gyatso, K., Norbu, N., Lodroe, T., ... Fresco, M. (2020). Inter-brain synchronization in the practice of tibetan monastic debate. *Mindfulness*, *11*, 1105-1119. doi: doi:10.1007/s12671-020-01338-1

van Vugt, M. K., Simen, P., Nystrom, L. E., Holmes, P., & Cohen, J. D. (2012). EEG oscillations reveal neural correlates of evidence accumulation. *Frontiers in Neuroscience*, *6*. doi: 10.3389/fnins.2012.00106

Verdiere, K. J., Dehais, F., & Roy, R. N. (2019). Spectral EEG-based classification for opera-

tor dyads' workload and cooperation level estimation. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *2019-october*, 3919-3924. doi: doi:10.1109/SMC.2019.8913848

Vézard, L., Legrand, P., Chavent, M., Faïta-Aïnseba, F., & Trujillo, L. (2015). EEG classification for the detection of mental states. *Applied Soft Computing Journal*, *32*, 113-131. doi: doi:10.1016/j.asoc.2015.03.028

Wei, Z., Wu, C., Wang, X., Supratak, A., Wang, P., & Guo, Y. (2018). Using support vector machine on EEG for advertisement impact assessment. *Frontiers in Neuroscience*, *12*, 76-88. doi: doi:10.3389/fnins.2018.00076

Zhao, H., Guo, X., Wang, M., Li, T., Pang, C., & Georgakopoulos, D. (2018). Analyze EEG signals with extreme learning machine based on pmis feature selection. *International Journal of Machine Learning and Cybernetics*, *9*, 243-249. doi: doi:10.1007/s13042-015-0378-x

Zwillinger, D., & Kokoska, S. (2000). *Crc standard probability and statistics tables and formulae*. New York: Chapman & Hall. doi: doi:10.1201/b16923

A Appendix

The appendix contains data that was too tedious to present in-line.

Data source	Shrinkage LDA	SVM
θ inter-brain synchrony	F(9,90) = 0.1461, p = 0.9981	F(9,90) = 0.2877, p = 0.9765
θ power challenger	F(9,90) = 0.4002, p = 0.9319	F(9,90) = 0.3405, p = 0.9589
θ power defender	F(9,90) = 0.1137 p = 0.9993	F(9,90) = 0.3423, p = 0.9583
α inter-brain synchrony	F(9,90) = 0.2404, p = 0.9875	F(9,90) = 0.3254, p = 0.9646
α power challenger	F(9,90) = 0.2670, p = 0.9819	F(9,90) = 0.1223, p = 0.9991
α power defender	F(9,90) = 0.1400, p = 0.9984	F(9,90) = 0.8800, p = 0.5462

Table A.1: One-way ANOVA statistics of the first feature importance analysis.

Data source	Shrinkage LDA	SVM
θ inter-brain synchrony	F(9,90) = 0.1467, p = 0.9981	F(9,90) = 0.5297, p = 0.8495
θ power challenger	F(9,90) = 0.5636, p = 0.8234	F(9,90) = 0.0506, p = 1
θ power defender	F(9,90) = 0.1355, p = 0.9986	F(9,90) = 0.2546, p = 0.9847
α inter-brain synchrony	F(9,90) = 0.2300, p = 0.9893	F(9,90) = 0.2696, p = 0.9812
α power challenger	F(9,90) = 0.3785, p = 0.9426	F(9,90) = 0.9630, p = 0.4758
α power defender	F(9,90) = 0.0804, p = 1	F(9,90) = 0.9630, 0.4758

Table A.2: One-way ANOVA statistics of the second analysis.