DECODING MOTOR INTENTIONS

CS 530 FINAL REPORT

Daniel Briseno Servin   Lucas Jeay-Bizot   Jeremy Wayland

May 23, 2021

1 Introduction

1.1 Background

Free Will and the Brain

As neuroscience's scientific enterprise flourishes, it becomes clearer that many cognitive processes follow a set of rules at the neural level. Many have argued that these advances indicate that our actions are fully determined by our neural events [2]. That is to say, that because neurons in our brain are ruled by deterministic processes, and because their firing determines our actions, our actions cannot be free. In our study, we aim to chirp a little bit at this bigger question and investigate whether information, at the electrical level, about an upcoming decision is available in the brain before that decision is made. The temporal relation between the encoding of the decision in the brain and the moment of the conscious awareness of that same decision is key here. Indeed, for neural determinism to be false, the decision should not be encoded prior to the moment that it is consciously made. On the other hand, if neural determinism is true, then it would very well be plausible that a decision is first constructed unconsciously at the neural level before being brought to awareness.

The Timing of Free Decisions

In 1965, a team of researchers found a brain signal, the Bereitschaftspotential, that precedes voluntary actions by a few seconds [4]. Years later, in 1983, another team of researchers found that participants report their awareness of their intention to move to be nearly half a second later than the onset of that signal [5]. Many have interpreted these results to mean that the decision is first made unconsciously and only later on misattributed by the agent to his or her own volition [5][3][2]. Taken as such, this result can be threatening to deterministic stances on the topic of free will. However, as it has been argued before, this decision is about "when to move" and not "what to move" [1]. The exact timing of a decision is not commonly perceived as being as important as the decision itself. In other words, if one were to show that information about a "what" decision is present in the brain prior to one making a decision, then that would represent a greater threat to the deterministic notion of free will then being able to predict when. Indeed, most moral choices are not about when to do something, but rather about what to do. For instance, when a judge chooses a prison sentence for a convict, in most cases, it matters more what the decision about the prison sentence is rather than when the judge forms that decision.

Predicting the Will

Recent works have investigated the predictability of a free "what" decision [7][6]. One such work uncovered, using functional magnetic resonance imaging, that a free left-right button press could be decoded from brain activity with nearly 60% accuracy up to 8 seconds before the movement occurred using multivariate pattern classification analysis [7]. Another work, using electro-encephalography (EEG), uncovered that a free left-right button press could be decoded with up to 75% AUC up to 0.5s before movement using a linear discriminant analysis approach [6]. However, in the latter study, the participants were instructed to make their decisions at a pre cue that preceded the go cue. In our project, we aim to investigate brain activity that precedes the conscious decision.

1.2 Research Question

In this project, we aim at addressing the question about how much information is present in the brain prior to a free "what" decision. To do so, we acquire electro-encephalographic data while participants are free to choose between a left and right button press. We ask participants to make spontaneous decisions between a left and right button. We hypothesized that the time course of decoding accuracy for a spontaneous free decision would be displaying lower accuracy and later onset than the time-course for an instructed movement. We did so because in the latter condition the information is already present at the conscious level in the participant while in the former it only reaches conscious awareness close to the movement onset.

2 Methods

2.1 Data Collection

The Task

In our study we invited 2 participants to come in and perform a button press task. They had their left index finger positioned above the 'q' key on the keyboard and the right index finger above the 'p' key. On each trial they had to keep their gaze fixated at a fixation cross while a white dot was running along an annulus (See Figure 1). On each trial a red target appeared on the annulus at a random position such that it took from 2 to 5 seconds for the white dot to reach it. Upon reaching the target, the participants pressed either the left or the right button. Which button they pressed depended on the condition:

- **Condition A** - *Instructed* : At the start of each trial, participants were instructed which button they should press during the upcoming trial.
- **Condition B** - *Spontaneous*: Participants were instructed to withhold any decision about which button to press until the white dot reached the target and then spontaneously press either the left or the right button.

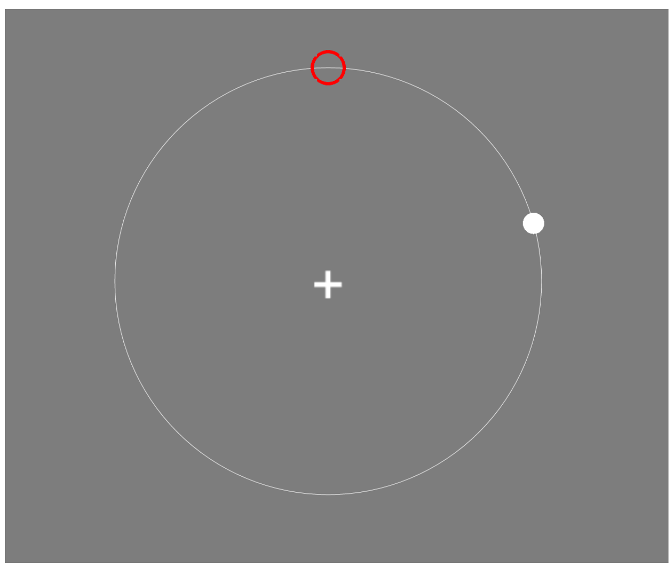

Figure 1: Trial, annulus with white dot rotating around clockwise at a pace of 4s per rotation and fixed red target.

By virtue of the instructions, it is an assumption of our study that the decision about which button to press in condition B only entered conscious awareness close to the movement. Participants were asked afterwards about the spontaneity of their decisions during the session to ensure that they adhered to the instructions. Both subjects indicated verbally that they had performed condition B correctly.

Equipment

Data was acquired using the BrainVision Recording software together with an actiCHamp amplifier and an EasyCap headcap with 64 channels. This resulted in 64 one dimensional signals of the scalp's electrical potential. The electrical signal is measured in volts at the order of microvolts. The data was recorded at a sampling rate of 1000Hz. To epoch the data, all trials data were truncated to match the length in samples to the shortest epoch. The data was then transformed into data matrices of three dimensions: channels (64), samples (subject 1: 3579; subject 2: 1937) and trials (subject 1: 200; subject 2: 452).

2.2 Pre-Processing

The data was pre-processed to remove noise and common EEG artifacts. Independent component analysis (ICA) was performed to remove eye-artifacts as we believed that it would greatly affect classification accuracy. Further, along a similar rationale we manually removed trials that contained muscle and channel artifacts in subject 2. Both of these analysis steps were performed using the FieldTrip Toolbox in MATLAB. Tentatively, we tried applying a low-pass filter (at 35 Hz and 70 Hz) to our data as we believed that high frequency noise would be irrelevant to our classification, however this proved to greatly decrease the accuracy (see Results section). We therefore settled for un-filtered data in our final analysis. Further to classify our data we used two dimensionality reduction techniques: common spatial patterns (CSP) and using a metric inspired by the lateralized readiness potential (LRP). CSP was performed using the MNE toolbox in python and LRP was performed by taking the difference in signal between channel C4 and C3.

We plotted the event related potential (ERP) at channel C4 (left motor cortex) to visualize the separability of the data. Due to the data being hardly separable (Figure 2) we hypothesized that the feature space might require reshaping for a classification task.

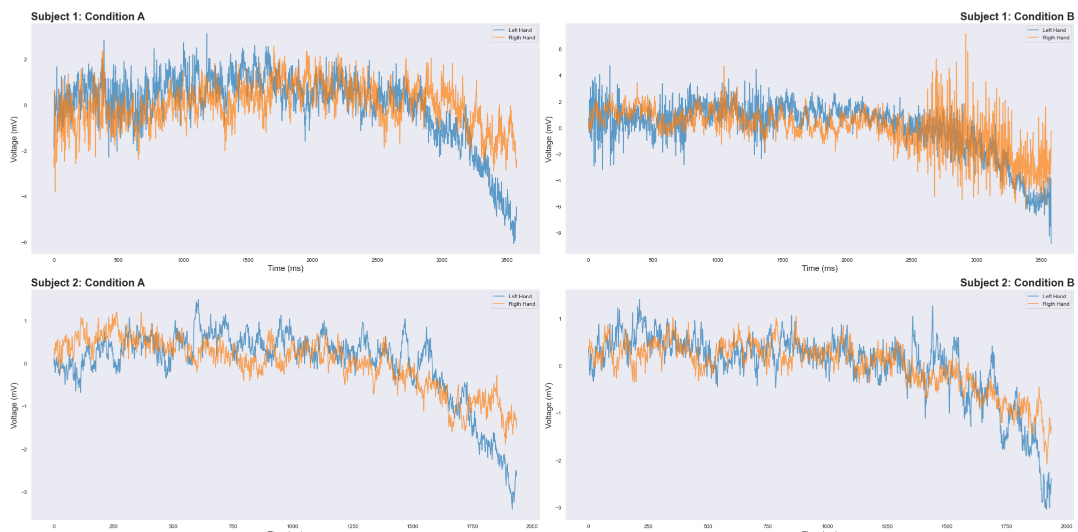

Figure 2: Seperability Plot

2.3 Machine Learning Models

To answer our research question we decided to first identify a best performing model for classification and then investigate the time course of that model's accuracy in the moments preceding a movement. We investigated four different models: linear discriminant analysis (LDA), support vector machine (SVM), random forest (RF) and AdaBoost (AB).

LDA is a machine learning algorithm that finds a linear combination of features to maximize the separability of the data. In our case, LDA seems a priori well suited for our data since as mentioned above we would wish to find a space to separate our data. SVM is an algorithm that also tries to separate the data, it does so by maximizing the gap between two classes in a linear separation. Since our data is non-separable, using kernels with SVM will most likely best fit our data. Kernels are operations on the data that reshapes the feature space to uncover new separability. RF is an ensemble learning algorithm that classifies by averaging the output of multiple decision trees. In RF, bagging is used together with a random subset of features used at each tree. This avoids overfitting and can deal well with separability issues too. AB also uses trees as weak learners with an emphasis put on misclassified instances at each new weak learner. The ensemble of these weak learners then converges to a strong learner. We believed that each of these algorithms was well suited to our problem as they would deal well with our small sample sizes. Our analysis was divided into two major axes: hyper-parameter tuning and sliding window accuracies.

Hyper-parameter Tuning

For hyper-parameter tuning, we ran a grid search through a breadth of parameters for each of our four models. The specific hyperparameters iterated over can be found in the code files. For each classifier type, we tuned two separate models, one tuned on data treated by MNE's CSP transform and another using the LRP approach to the data.

To avoid double-dipping with the CSP transform, CSP was fit on 40 trials of data using the default number of components, then the remaining trials were used for hyperparameter tuning. Note that we did not exclude any additional trials from the hyperparameter tuning, as would be traditionally done to avoid overfitting. This is for the following reasons:

- For this analysis, we are interested in observing patterns in data that we already have, thus, we are not worried about our validation results extending to unseen data.
- No matter how well we choose the hyperparameters, a classifier will not be able to "memorize" the classification unless it is trained on the testing data. Thus, this hyperparameter tuning will not allow us to see classification separability that is not present in the data.
- The dataset has a small amount of trials compared to the dimensionality of the data. Thus, we would like to tune the hyperparameters on as much of the data as possible.

The above considerations do not apply to fitting the CSP transform, since fitting CSP uses the target labels to transform the data. Thus, fitting CSP to the entire dataset before tuning would give additional correlation between features and labels which will not exist during the sliding-window phase of the analysis, and so non-optimal hyperparameters might be chosen.

After fitting the hyperparameters for each classifier, we now re-chose the n_components hyperparameter to CSP which sets the remaining number of features in the data after the CSP transform. This parameter was chosen from a range [1,63] using 5-fold cross validation.

Sliding-window Analysis

For the sliding window we generated a model and its prediction for windows of different width (100, 200, 300, 500 and 1000ms) and settled on 500ms as the best window width. We generated accuracy time courses by generating models and testing them in windows of width 500ms and sliding them over the epochs by steps of 10 samples. The results from the different windows timing can be found in the github repository under the Plots folder. Due to the very high run-time of hyperparameter tuning (see Table 1) we opted to generate parameters outside of our sliding window. Models were thus all hyper-parameter tuned globally on the whole epochs of the entire dataset while the models were generated locally within the train set of each local window.

	Random Forest/LRP	AdaBoost/LRP	LDA/LRP	SVM/LRP
Hyperparameter Tuning Time	1hr 19min 19s	0hr 31min 54s	0hr 1 min 47s	0hr 1min 49s

Table 1: Time taken to tune hyper-parameters for several classifiers using LRP.

3 Results

3.1 Model Tuning

After running hyper-parameter tuning on both subject 1 and subject 2 data. The accuracy plots can be seen below (Figures 3, 4, 5, and 6). They clearly indicate that CSP performs poorly, as such we decided to drop it from further analysis. We believe that this might be due to our dataset being too small for 64 dimensions of data (64 channels). By doing feature selection and keeping only a simple linear combination of 2 channels in LRP we believe that this circumvents the curse of dimensionality that leads to small accuracies when using CSP.

For subject 1, we see that for classifiers using LDA the accuracies are comparable. Thus we plotted distributions of accuracies resulting from random test-train splits with $n = 200$. From this, we see that Random Forest and LDA were the most reliable, since they both had the least amount of trial accuracies near the random accuracy mark.

For subject 2, all models seemed to perform on an equal level with SVM taking the lead. This suggests that with a greater dataset all models are suited for the task at hand. However, looking at distribution plots, we can see that LDA and RF remain the most reliable (least variance). As a result, for subject 2 we decided to also plot a sliding window for RF and LDA. For both subjects we restricted our sliding window analysis to the LRP pre-processed data.

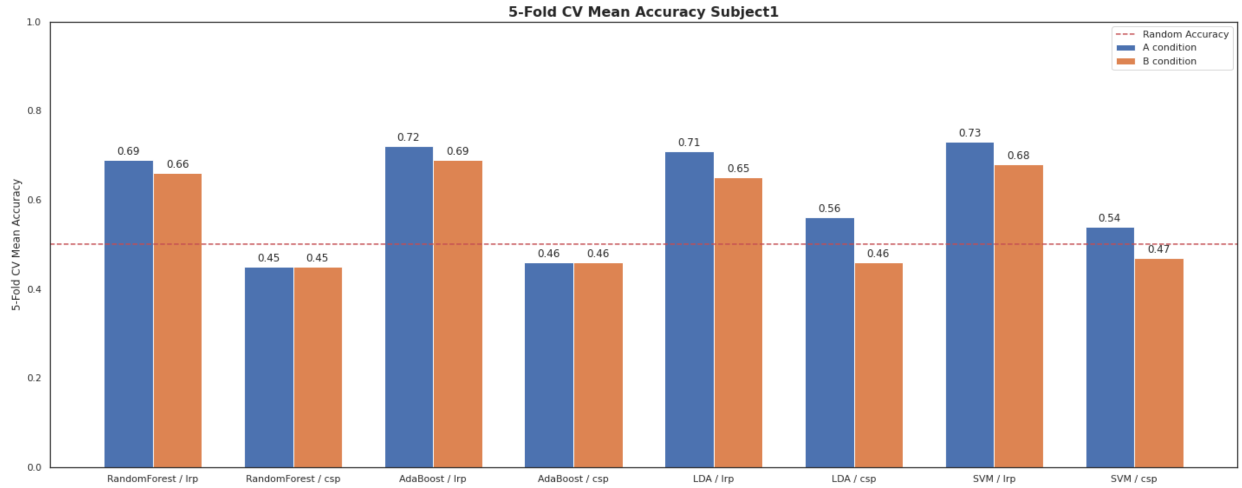

Figure 3: Cross-validation best hyper-parameter tuned model's accuracy (Subject 1)

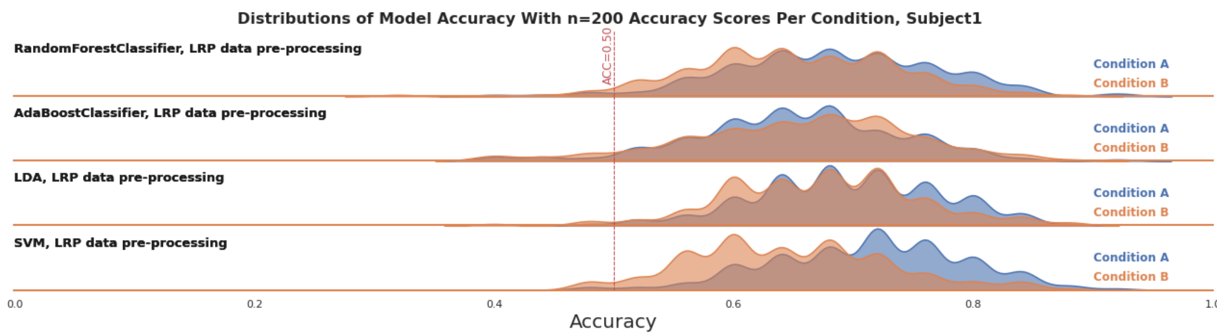

Figure 4: Distribution of accuracies for LRP best tuned models (Subject 1)

3.2 Sliding Window Predictions

Our results indicate that there is no difference in the time-course of the accuracy between both conditions (See Figure 7 and 8). From the time course of the sliding windows in subject 1 we can see that the decoding accuracy picks up about a half second before movement. This corroborates previous results in the literature [6]. For subject 2, the decoding accuracy is above chance during the entire epoch, and there seems to be no distinctive dynamic pattern nor difference between the two conditions. This is harder to interpret, however due to the smaller epochs in subject 2 (due to smaller clock rotations during data acquisition) it is hard to interpret this above chance accuracy. On the other hand, it is obvious that both conditions' decoding time course does not seem to evidently differ neither over time nor overall. This goes against our initial hypothesis and further goes against the ERP results (Figure 2). Indeed, although the signal over the motor cortex behaves differently across the two conditions, the decoding accuracy of our classifier does not. Our classifier, in the LRP models, used only information from channel C3 and C4. This indicates that despite the two signals looking different to the naked eye, both bear enough information about the upcoming movement so that their decoding accuracy can be indistinguishable. We had hypothesized that in condition B, the spontaneous condition, due to the non-awareness of the upcoming motor action in the participant, decoding accuracy would differ from condition A, the instructed condition. However, our results show that this is not the case (in our sample). This result indicates that information about an upcoming decision is present in the brain before one is aware of it, or at least, that in a free spontaneous decision the amount of information does not differ from that present in an instructed condition.

3.3 What Went Wrong

During this project we came across many difficulties and learned about what works and what does not work when applying machine learning algorithms to neural data.

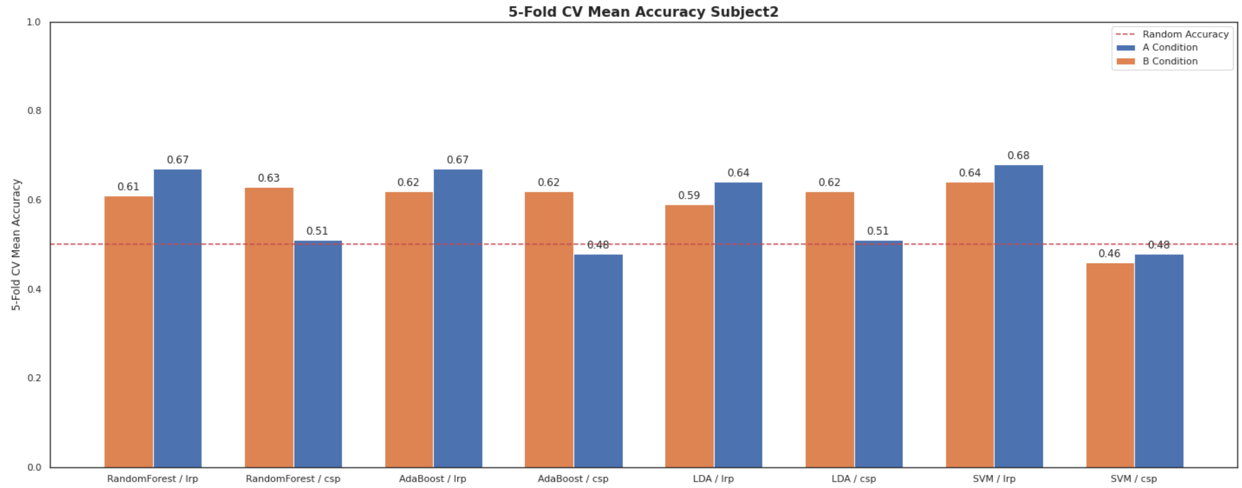

Figure 5: Cross-validation best hyper-parameter tuned model's accuracy (Subject 2)

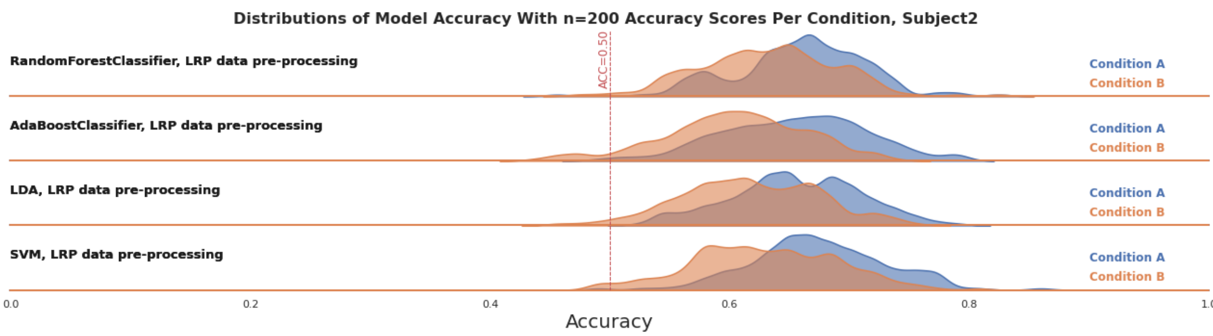

Figure 6: Distribution of accuracies for LRP best tuned models (Subject 2)

Firstly, hyper-parameter tuning proved to have enormous run-times for RF and AB. We had initially intended to hyperparameter tune our classifiers on each sliding window, but these run times forced us to tune our hyper-parameters globally (that is, without taking sliding windows). Moreover, SVM failed to converge for data pre-processed by CSP. Thus, to avoid infinite runtime, we needed to set a maximum number of iterations to 1000. This meant that for data pre-processed by CSP, we used and evaluated a badly fit SVM. This did not impact our final analysis, since we rejected CSP as a data pre-processing method.

Elaborating on the poor performance of CSP, CSP generally did not work with any of our tested classifiers, even though it is a method specifically designed for EEG analysis. Using CSP resulted in accuracies that were no better than random. We believe that this might be due to our small sample sizes. We therefore had to resort to feature selection instead and selected only a linear combination of two features (C3 and C4).

In our pre-processing steps we had also attempted to remove noise from our data in order to increase separability by applying a low-pass filter. However, similarly to CSP, using low-pass filters (35Hz and 70Hz) both resulted in accuracies that did not differ greatly from chance. We therefore decided to keep our data unfiltered. This might be due to the fact that information relevant for classification is present in these very high frequencies.

Plots generated from the above failed attempts to pre-process using CSP and low-pass filtering can be found in the "Plots" folder in the repository.

4 Conclusion

Overall, the striking similarity between the decoding accuracies' time-courses of the two conditions can indicate one of many things. Firstly, it is possible that we committed a mistake in our analysis. However, when re-running it step by step on the second subject 2 we produced similar results. Secondly, it is possible that our participants did not perform

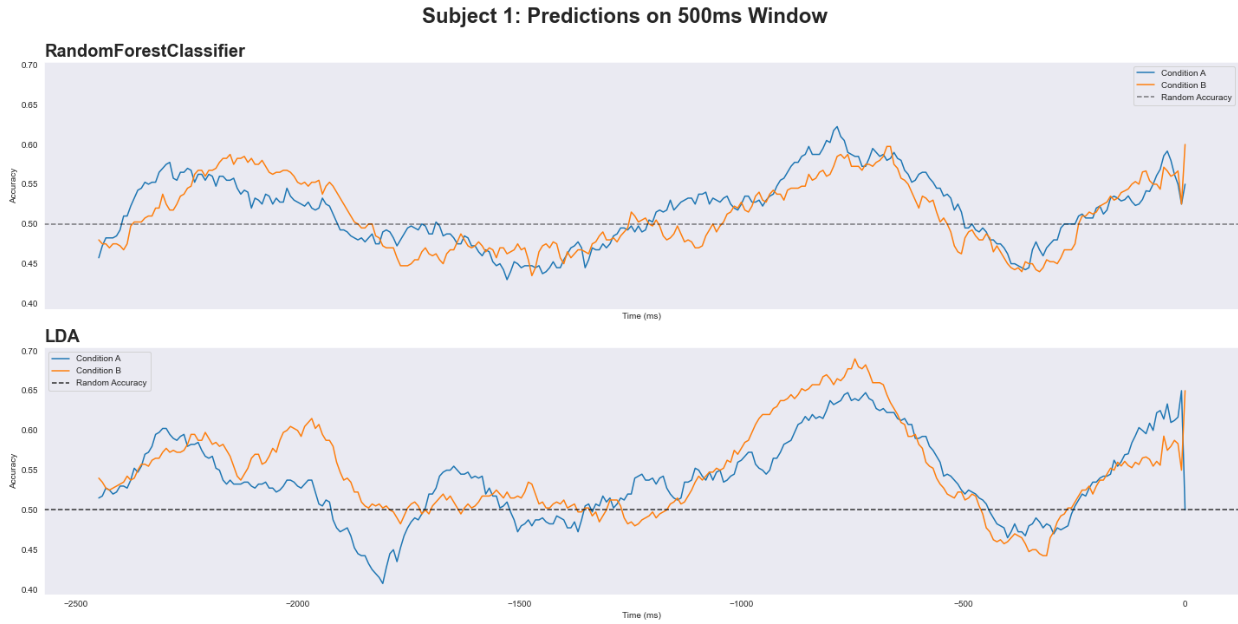

Figure 7: Classification accuracy with sliding window of 500ms width (Subject 1)

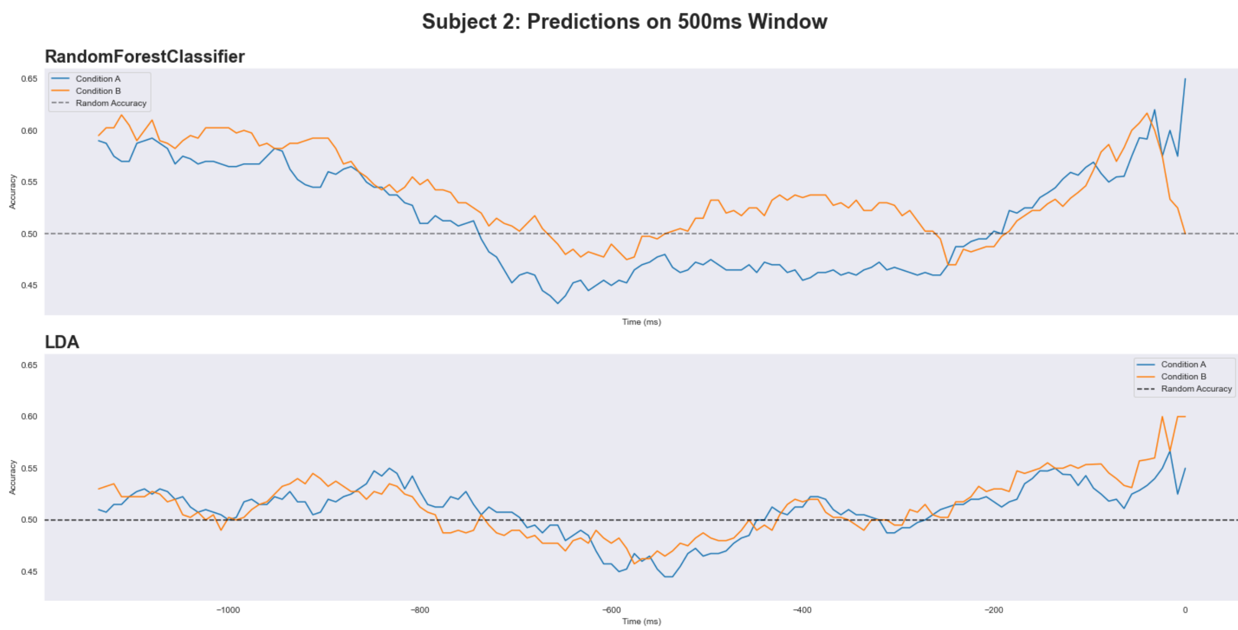

Figure 8: Classification accuracy with sliding window of 500ms width (Subject 2)

the task in condition B correctly. However, we checked with them after the task verbally and their report indicated that they had performed the task accordingly. Finally, it is possible that both conditions' decoding accuracy are similar because the information used for decoding the intention is independent from the participant's awareness of his or her decision. If reproduced at a larger sample size, this is an interesting result as it would indicate that before we are aware of our decisions, our brain already encodes the content of that decision.

References

[1] Marcel Brass and Patrick Haggard. The what, when, whether model of intentional action. *Neuroscientist*, 14(4):319–325, 2008.

[2] Anthony R. Cashmore. The Lucretian swerve: The biological basis of human behavior and the criminal justice system. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4499–4504, 2010.

[3] Joshua Greene and Jonathan Cohen. For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359(1451):1775–1785, 2004.

[4] Hans Kornhuber and Lüder Deecke. Hirnpotentialänderungen beim Menschen vor und nach Willkürbewegungen , dargestellt mit Magnetband-Speicherung und Rückwärtsanalyse Ilirnpotential ∼ nderungen bei Willkarbewegungen und passiven Bewegungen des Mensehen : Bereitschaftspotential und reafferen. *Pflügers Archiv*, 284(1):1–17, 1965.

[5] Benjamin Libet, Curtis A Gleason, Elwood W Wright, and Dennis K Pearl. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106:623–642, 1983.

[6] Mathew Salvaris and Patrick Haggard. Decoding intention at sensorimotor timescales. *PLoS ONE*, 9(2):1–11, 2014.

[7] Chun Siong Soon, Marcel Brass, Hans Jochen Heinze, and John Dylan Haynes. Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 11(5):543–545, 2008.

5 What each of us did

- **Lucas**: Data collection, data pre-processing, LDA, introduction and conclusion write up.
- **Daniel**: Hyper-parameter tuning, LDA, SVM, RF and AB, plot generation/figures.
- **Jeremy**: Sliding-window analysis, SVM and plot generation/figures.
- **Common**: Planning the analysis, writing the methods and results, weekly meetings, discussing the results.