Determining the nature of free will using machine learning

by

Siobhan Hall

Thesis presented in partial fulfilment of the requirements for the degree of Master of Physiotherapy in the Faculty of Medical and Health Sciences at Stellenbosch University

Supervisors:

Dr L.D. Morris

Prof D. van den Heever

March 2020

Faculty of Medicine and Health Sciences

Division of Physiotherapy

Private Bag X1, Matieland, 7602, / South Africa · Francie Van Zijl Dr, Cape Town, 7505, South Africa

Tel: +27 21 938 9300 / Email: postgradphysio@sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Name: Siobhan Hall

Date: March 2020

Abstract

Background

The debate around free will has been topical for millennia. The question around free will is important in assigning agency to our decisions and actions. The definition of free will used in this research is the ability for a person to do otherwise, should the exact circumstances be created. In 1983, the Libet paradigm was developed as a means to empirically investigate the nature of free will. The Libet paradigm resulted in the presentation of a rise in neural activity 350 ms before conscious awareness of a decision to act. This rise in neural activity (known as the readiness potential) was prematurely and incorrectly taken as proof that the subconscious having a prominent role in our decision-making processes and therefore the conscious self has no free will. This result has subsequently faced criticism, particularly its method of averaging out EEG data over all the trials and the readiness potential not being present on an individual trial basis. Another major criticism is the method of retrospectively and subjectively reporting the moment of conscious awareness, termed “W”.

Objectives

The aim of this research is to determine the role of the subconscious in our decision-making processes using machine learning. A secondary aim is to determine if eye tracking can be used to objectively mark the moment of conscious awareness of a decision to move. Investigating the role of the subconscious in our decision-making processes not only contributes to the fundamental understanding of our brains’ processes and the nature of free will, but also early detection of intentions to move can aid in the earlier identification of features to classify actions in brain-computer interface (BCI) systems. This earlier classification can improve the real-time nature of thought and then action. This can help improve the functionality of people living with disabilities.

Methodology

The data collection involved the recreation of the Libet experiment, with electroencephalography (EEG) data being collected in conjunction with eye tracking. Another addition to the Libet paradigm was the choice between “left” and “right”. 21 participants were included (4 females, all right-handed). The participants were asked to make a decision between moving “left” and moving “right” while observing the Libet clock to subjectively mark the moment of subconscious awareness. Deep learning, a branch of machine learning was used for the EEG data analysis. The deep learning model used is known as a convolutional neural network (CNN). The eye tracking data was used to identify any eye movements (saccades) that occurred 500 ms before the action.

Results

The CNN model was able to predict the decision “left” or “right” as early as 1.3 seconds before the action with a test accuracy of 99%. The eye tracking data was analysed and no correlations between an eye movement and the moment of conscious awareness was found.

Conclusion

This research has provided evidence to support the hypothesis that there is no free will. Further research is needed to investigate earlier predictions using deep learning as well as research focused on using eye tracking as a means to objectively time-lock the moment of conscious awareness.

Opsomming

Agtergrond

Die kwessie van vrye wil is duisende jaar oud. Dit is belangrik om bemiddeling aan ons besluite te gee. In hierdie navorsing, beteken vrye wil om iets anders te kan doen of ’n in presies dieselfde omstandighede alternatiewe aksies uit te voer, sou die persoon of persone so besluit. Libet en sy span het in 1983 ’n eksperiment ontwikkel om die kwessie van vrye wil te toets. Hulle het uitgevind dat daar 350 ms voor die persoon bewus geword het van hulle besluit om te beweeg, reeds breinaktiwiteit plaasgevind het. Hierdie aktiwiteit is die ‘readiness potential’, of beruitschaft-potential’ genoem. Libet het tot die verkeerde slotsom gekom dat die ‘readiness potential’ bewys dat ons besluitneming in die onderbewussyn van ons brein begin en as gevolg daarvan is daar *geen* vrye wil nie. Onlangse navorsing het hierdie resultaat gekritiseer omdat die ‘readiness potential’ ’n gemiddelde teken van breinaktiwiteit is, en is nie beskikbaar voor enkele besluite nie. ’n Ander belangrike kritiek hieroor is oor die subjektiewe wyse waarop bepaal word hoe om die oomblik van bewustheid van ’n besluit te meet.

Doelstellings

Die primêre doel van die navorsing is om die rol van die onderbewussyn in ons besluitnemingsproses te verstaan. Die sekondêre doel is om oogbeweging te gebruik om die oomblik van bewustheid van ’n besluit te meet. Hierdie navorsing sal help met die fundamentele begrip van die brein se prosesse en vrye wil. Om hierdie vroeë vasstelling van besluite te kan maak sal ook help met die ontwikkeling van brein-rekenaar-interaksie sisteme. Dit sal help om meer tydelike en natuurlike bewegings te ontwikkel. Dit sal die funksionele potensiaal van mense met beserings beïnvloed.

Metodes

Die data-insamelling is gebaseer op die oorspronlike Libet-eksperiment. Daar was twee verskillende data-tipes wat opgeneem is: elektroënselografie (EEG) en oogbewegings. ’n Verandering aan die oorspronklike Libet eksperiment was gemaak: die deelnemers moes tussen “links” en “regs” kies. Een-en twintig regshandige deelnemers is ingesluit van wie vier vrouens was. Die deelnemers het hulle keuses gemaak terwyl hulle die Libet horlosie dopgehou het. Die Libet-horlosie word gebruik om die oomblik van bewustheid van ’n besluit te meet. Masjien-leer algoritmes is gebruik om die EEG data te analiseer. ’n Verwikkelde neurale netwerk is gebruik. Die oogbewegingsdata is geanaliseer om oogbeweging 500 ms voor die aksie te probeer identifiseer.

Resultate

Die verwikkelde neurale netwerk kon die besluit “links” of “regs” met 99 % akkuraatheid voorspel. Die voorspelling is 1.3 sekondes voor die aksie gemaak. Daar was geen korrelasie tussen die oogbeweging en die oomblik waarop die besluit bewustelik gemaak is nie.

Slotsom

Die navorsing het meer bewyse gelewer ten opsigte van die hipotese dat daar geen vrywilligheid is nie. Verdere navorsing is nodig om vroeëre voorspellings met masjienleer-algoritmes te kan maak en nog meer navorsing is nodig om die korrelasie tussen oogbeweging en die oomblik van besluitneming bewustheid te verstaan.

Acknowledgements

Significant contributions were made to this work by the following people: Professor Dawie van den Heever, Dr Linzette Morris, Quentin Hall, Elan van Biljon, Stuart Reid, Joshua Fischer, Marisa Coetzee, Dr Mikkel Vinding, Guillaume Odendaal, Benjamin Wolfaardt, Andreas van der Merwe, Prof Daan Nel, Riëtte Hugo, Leonard Botha, Dr Arnaud Klopfenstein and Dr Thazin Htwe. Acknowledgements and thanks are also given to Professor Jochen Baumeister and Daniel Büchel of the Neuroscience and Exercise Unit at Paderborn University, Germany for their assistance in teaching EEG data collection and pre-processing.

Very special thanks are given to the Biomedical Engineering Research Group of Stellenbosch University for undertaking and supporting this collaborative project.

Table of contents

Page

List of figures

Page

List of tables

Page

Clarification of concepts

Concept	Explanation
Conscious(ness)	These are the thoughts and cognitive processes of which we are aware. This allows for interaction with our environment
Subconscious(ness)	In this research the distinction is made between unconscious and subconscious to avoid confusion. Subconscious refers the underlying brain function that occurs while one is in the state of being awake. This is the brain function under the hood of the *conscious* brain function
Unconscious	This is the state of not being awake and not being engaged with our environment
Free will	Agency is assigned to the conscious-self, meaning each person is completely in control of their own actions, without any external influences. There is also the ability to have done otherwise, or do otherwise should the exact same circumstances be recreated
A lack of free will	In the context of this research, a lack free will is defined as the subconscious having control over our actions. The consciousness merely witnesses the product of the subconsciousness's actions and decisions. Agency is assigned to the subconscious and the conscious-self has no control over it: the conscious is not able to do otherwise
EEG	Electroencephalography This is the non-invasive recording of the electrical potentials of the brain from the scalp
ICA	Independent component analysis Used in the pre-processing steps of EEG data preparation.
ERP	Event related potential
RP	The readiness potential found in the original Libet paradigm. This example of an ERP
EMG	Electromyography This is the non-invasive recording of the electrical potentials of the muscles
ECG	Electrocardiography This is the non-invasive recording of the electrical potentials of the heart and the associated arterial and venous pulses

ET	Eye tracking
SMA	Supplementary motor area
Pre-SMA	Pre-supplementary motor area
AAC	Anterior cingulate cortex
"M"	This is the moment the action takes place, in the Libet paradigm
"W"	This is the moment of conscious awareness of an intention to move
Trial	This refers to a single Libet clock – from the start of the clock to the moment of action
Round	This refers to a collection of 11 consecutive trials
	A collection of trials is referred to as an experiment (for the party involved)
Experiment	The entire data collection period for one participant
Eye-time	This refers to the time-locking of eye events in the eye tracking data
Model	This refers to the machine learning (or, more specifically, deep learning) algorithm employed in order to complete a task. The term is used interchangeably with algorithm in the context of machine learning, as well as the term neural network in the context of machine- and deep learning
SLA	Supervised learning algorithm
DL	Deep learning
CNN	Convolutional Neural Network
Hyperparameters	Selected by the programmer
Parameters	Learnt by the network, includes weights and bias
Weights	Transform information as it flows forwards through the network

Windows	This is the EEG matrix segmented for analysis Formally known as an epoch in neuroscience, but will be referred to as a window to avoid confusion with a training epoch in machine learning.
Epochs	In machine learning, an epoch refers to the process of a single forward propagation and backward propagation. The model trains on the training set, outputs a training accuracy. It then tests its performance on the validation set. The model then compares the training and validation accuracies and updates its weights accordingly.
Frames	This is the percentage of the window of EEG data fed as input into the model
BCI	Brain computer interface This refers to the type of prosthetic (or similar) controlled by monitoring the person's electrical potentials (EEG) and outputting an action
PLWD	Person(s) living with disabilities
ADLs	Activities of daily living

List of symbols

$h_{(\theta)}$	Hypothesis function parameterised by θ
θ	Parameters learned by the machine learning model
y	Label provided to the MLA which refers to the "actual answer"
x	Input training examples to the machine learning
$J(\theta)$	Cost function

1 Introduction

The following chapter introduces the topic of the thesis, the aims and the rationale for the thesis.

1.1 Background

I have noticed, even people who claim everything is pre-destined, and that we can do nothing to change it, look before they cross the road.

Stephen Hawking

The question of free will has been topical for millennia, especially considering its links to moral responsibility and the ownership of that responsibility. Who, or what, is accountable for our thoughts and actions? Who, or what, is pulling the strings? Who is ultimately in control; and who, or what, has the ability to make our choices (Burns & Bechara, 2007)? Is it the self, or an unknown determining entity or simply the consequence of nature and its determining laws? This question ultimately seeks to give, or take away the agency, or ownership, to the self and to our decision-making processes.

Furthermore, what constitutes free will and its subsequent ability to control our actions? What exactly drives our volition and the ability for us to choose for ourselves? A free choice requires options, to avoid being denounced as the result of 'unspecific neural preparatory action" in the brain (Soon *et al*., 2008). There is the specific requirement of being able to do otherwise, should one choose to (Dias & Lavazza, 2016). The alternative to this is a completely random, or irrational decision, which would typically present as a reaction or urge, and not an act of volition. A reaction, generally based in instinct, cannot quantify as a volitional action. These neural pathways are ingrained following millennia of self-preservation driven evolution. These free choices also need to be carried out in the real world, in real world situations, with real world consequences, but, also most importantly the decision needs to be bereft of any external or, internal processes that the person themselves is not in direct control over (Maoz *et al*., 2017).

The following sections will look in detail around these questions, initially focusing on the more philosophical viewpoints of free will, then progressing into the neuroscientific side of it, looking at the empirical evidence that has been collected over the years.

1.1.1 Are we free to decide?

… Freedom is always a question of degree rather than an absolute good that we do or do not possess.

Christof Koch

There are varying opinions on what determines, or allows, for our actions to occur. The two extremes in the debate take on the form of complete determinism and complete libertarianism. Determinism denotes the belief that everything that happens is governed by the laws of physics and is pre-determined by the events preceding them, with these subsequently determining events originating with the Big Bang. In other words, every action and reaction is set by *"the initial conditions of the universe [and] the laws of physics"* (Kuhn, 2014). The theory of reductionism stems from this, suggesting that if we were to "scientifically reduce" our "spiritual [self]" to components of biology, chemistry and physics, there would be no difference between the "laws governing [our] mind and the leaf blowing in the wind"(Perlovsky, 2011). French philosopher, Pierre Simon Laplace presented a thought experiment in 1814, now referred to as the "Laplace Demon":

> *We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes.*
>
> Pierre Simon Laplace, A Philosophical Essay on Probabilities

Essentially, what Laplace was suggesting is that were there an "intellect" (i.e. the so-called "demon") in this universe, with both knowledge of every causal and consequential factor that led to a specific moment and time, and the resultant effects and consequences of said moment; as well the means to analyse these factors, the intellect would know the future, as well as he knows the past (Laplace, 1902). To summarise the basis of determinism, I present the following comment on the "Laplace demon": "The present moment [is simply] the effect of its past and the cause of its future" (WikiAudio, 2016).

At the other end of the spectrum, libertarianism suggests that we are not the product of physical laws and their effect on events, but rather that the physical events occurring in the universe are

the product, or consequent result of our will and how we choose to act. Our conscious self is responsible for our actions (Vargas, 2004). The choice to think or act in a certain way is purely our own and were we to decide another course of action is more appropriate given the same circumstance, this option would be open to us – i.e. we are not the mere consequence of "neural processes in our brain" (Gomes, 2007), but, rather volitional beings. According to Wolpe & Rowe (2014) This gives specific "agency" or ownership to the "self" of a person, i.e. the conscious experience that one has control over their own actions. Through the exertion of this control, we as a person can affect the environment as well as be able to choose another course of action should we wish. Another way to explain this would be that a person who wishes to complete an "action A at time X", is free to choose "action B, under the same circumstance" should they have "willed" themselves to do so (Bode *et al.*, 2014). In essence, the physical laws of nature have no command over our will, but rather they bend to it.

In the middle of these two extremes, we can find other theories; each with a slightly different take on the idea of agency, or authorship of actions. Compatibilism somewhat merges the two extremes, claiming that some ideas are freely chosen, while others are automatic – free will takes on the form of a spectrum, with some ideas being the product of our 'true' self, while others are the product of consequence. In other words, the consequence of the laws of physics and nature. It accepts the notion that "natural phenomena are caused by other natural phenomena", and so a causal chain is created – and there's no reason to suggest that freely willed "conscious intention" would break this chain, but rather become a contributing and determining factor in itself. This notion accepts the compatibility of "freedom and natural causality" in terms of our actions. (Gomes, 2007)

Randomness takes the stance that things are exactly as they should be, however, there is no causing or determining factor. There is no causation chain pre-determining what happens, however, there is also no way to change or prevent an action from taking place. The present event is not a product of a chain of events originating in the Big Bang, nor is it a product of a freely willed decision. Christof Koch sums this up as, nothing could predict or determine an event [based on past events], but no-one had control over it either. (Kuhn & Koch, 2014)

The field of neuroscience has also tried to provide empirical evidence to help answer these questions. This will be discussed next.

1.1.2 Measuring (sub)conscious intent

In 1983, Benjamin Libet performed an experiment, which set out to question whether or not we have free will. Following on from the work of Kornhuber and Deecke, who discovered the so called "beruitschaft-potential" in 1965, otherwise known as the "readiness potential"(RP); Libet designed an experiment to study this sub-conscious neural precursor (Kornhuber & Deecke, 1965). This "readiness potential" is a neural precursor to movement and is found in the averaging of electroencephalography (EEG) data across many trials. The RP presents itself in the EEG data before the conscious awareness around a voluntary action takes place. The seminal work of Benjamin Libet set out to determine whether or not this readiness potential occurs before or after the person is consciously aware of their intention to act. (Libet *et al*., 1983)

The experiment comprised of 5 participants sitting in front of a clock, 1.95m away. The participants were asked to relax and fixate their gaze on this clock, which consisted of a cathode taking 2.56s to complete a revolution. The standard Libet clock is presented in Figure 1:

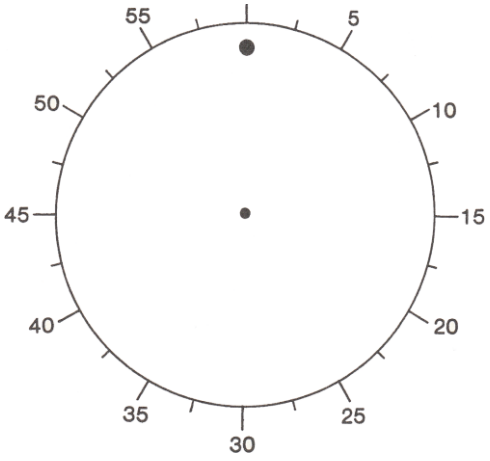


Figure 1: Image of the standard Libet Clock (Doyle, n.d.)

The participants were asked to lift either their left or right hands, once they felt the spontaneous "urge to move". Once the participant had moved, they were asked to report the moment (i.e. the corresponding position of the cathode on the clock face) that they became aware of this "intention" or "urge" to move. This is known as the "W" moment – the moment the participant is aware they want to move. The moment of movement (i.e. lifting either hand in this case) is also referred to as the "M" moment. (Libet *et al.*, 1983). This set up has since been used as a basis in many experiments, hence referred to as the 'Libet paradigm'.

The results of this study found the RP to be present 550 ms before actual movement ("M"), while the onset of the conscious awareness of the intention to act ("W") was found to occur 200 ms before movement ("M"). This means, there is a 350 ms period between the RP, the neural indicator for preparation of movement, and the moment of conscious awareness ("W") – i.e. 350 ms period of brain activity without our conscious awareness, or involvement. (Libet *et al.*, 1983).

Figure 2 depicts the RP and its corresponding time stamps as discussed above:

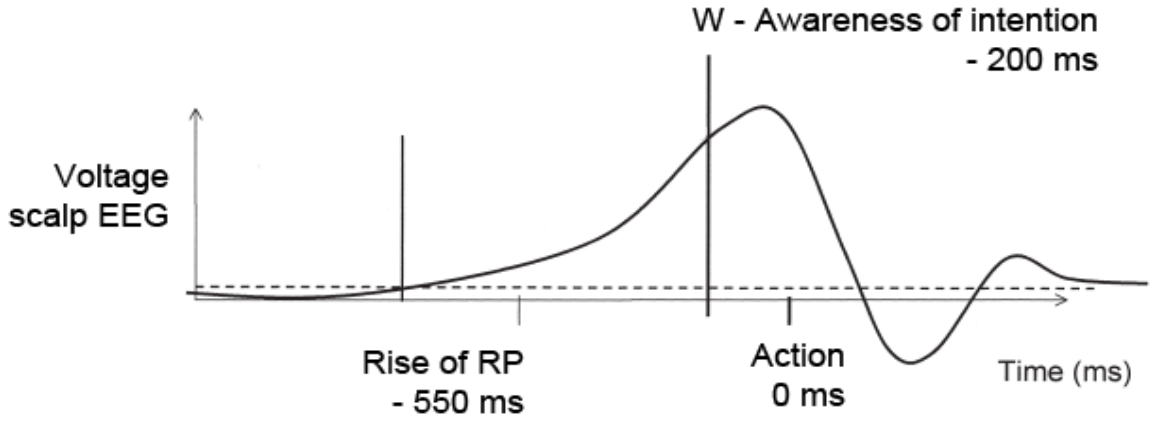


Figure 2: The time sequence of the readiness potential and the moment of conscious awareness (Doyle, n.d.)

This brings us to the question of free will – what is the brain doing on a sub-conscious level? Can we really be free, if the brain is acting out of its own volition, without our conscious input or awareness? The following studies attempted to shed further light on this.

According to Eagleman (2004), "There is a part of the brain moving towards a decision, before we are aware of it". The RP is further described by Eagleman (2004) as a "progressive rise in motor area activity prior to voluntary movement". The readiness potential is found in the supplementary motor area (SMA), the pre-supplementary motor area (pre-SMA) and the anterior cingulate cortex (AAC) (Eagleman, 2004). Haggard (2011) explains the RP to be a rise in motor activity, with the initial activity occurring on a subconscious level. It is only once this neural activity has reached a certain "threshold" that this intention to move enters conscious awareness. This reiterates the point that volitional action starts on a subconscious level. (Haggard, 2011)

Monitoring these early signals of "intent" (Yang *et al.*, 2015), and if possible even *earlier* predictive signals of intent, could lead to improvements in brain-computer interface (BCI)

prosthetics and the naturality of the movements produced. This potential impact, which goes beyond the philosophical debate, will be discussed in detail in Chapter 2.

In 2008, Soon *et al.*, sought to determine the areas of the brain responsible for determining movement on an subconscious level, and how early this process begins. Instead of the traditional Libet paradigm, the researchers had the participants watch a stream of letters in order to mark the moment of awareness of intent to click a button with either their left or right index fingers ("W"). Using functional magnetic resonance imaging (fMRI), as opposed to the standard EEG, the researchers found these decisions to be formed in the pre-frontal and parietal cortexes. They were able to predict which finger would move up to *seven* seconds before it entered the participants' conscious awareness.

In 2011, Fried *et al.* used a more invasive technique (depth electrodes) to study neurons during "self-initiated movements". The study focused on neuronal activity in the SMA, pre-SMA and the AAC as well as 259 neurons in the temporal region. The participants followed the Libet paradigm, the only difference being the participants were asked to press a key when they felt the urge to do so. "Progressive neuronal activity" was observed 1500 ms before the participant's report of making the decision to move, and the researchers could predict the movement 700 ms before this moment of decision ("W"). They concluded monitoring of only 256 neurons in the supplementary motor area is necessary to make this prediction.

Robert Lawrence Kuhn raised the question, "…How can we possibly have free will", if the "sense of authorship", or agency, (i.e. the sense of who or what the decision belongs to), only comes after the brain has decided what it will do (Kuhn & Koch, 2014)? Our brains make up our reality, deciding what is important information and what is not. In doing so, our consciousness jumps from one "logical island" to the next, creating our reality as we go along; ignorant to what the brain has decided is not important, or pertinent to existence at that moment (Eagleman, 2011). According to Harris (2012), "[We] are not in control of [our] minds…you, as a conscious agent are only a part of your mind, living at the mercy of other parts." The above experiments and statements attempt to understand and explain what our brains are doing on a subconscious level. The following addresses the issues with the Libet paradigm experiments as well as the brain's (subconscious) creation of our reality.

1.2 Problem statement

Much of current research into the question of free will is mired in criticism of methodology, especially in terms of the techniques of data collection and the analysis thereof. Neuroscientific research looking into the question of free will has mostly maintained tunnel vision in terms of its methods of data analysis, and as a result they have come under scrutiny and consequent criticism. This will be elaborated upon in chapter 2. This tunnel vision limits the analysis process – EEG has *millisecond* temporal resolution – this results in many innumerable time-points and features to be accounted for – and these purely neuroscientific methods end up excluding massive portions of this data during analysis (Johannesen *et al.*, 2016). Simple visual analysis of EEG data (as an applicable example), by a human observer introduces observer bias. This bias is not necessarily intentional, but, rather the product of the examiner's education, experience and preconceptions about what *should* be present in the data and where certain processes *should* take place in the brain. This bias can make manual interpretation of the data slightly subjective – however, this does not hold true for a computer: machine learning algorithms do not take into account what we as humans know and won't come with any pre-conceptions about how the brain should operate.

In terms of data collection, the limitations are as follows: the current means to time-lock "W", (the moment of conscious awareness), is to ask the participant for a post-hoc subjective report. Reporting the time one became aware of a decision requires a great deal of introspection and concentration – becoming a cognitively challenging task (Jo *et al.*, 2015; Schmidt *et al.*, 2016) that can detract from the aim of the Libet paradigm, which is to look at pre-conscious decisions and not focus on what happens during conscious awareness. Retrospective reporting has also been shown in the literature to be vulnerable to manipulation (Banks & Isham, 2008). There is a need for a more reliable and objective measure of "W". Evidence has shown the eyes correlate to attention and neural activity (Blignaut, 2009; Einhäuser *et al.*, 2010; Salvucci & Goldberg, 2000; Wierda *et al.*, 2012), creating a possible avenue to the "window to the soul"; the so-called adage, popularised by William Shakespeare. This presents an opportunity to objectively time-lock the moment of conscious awareness using eye tracking.

Current research, with the exception of Soon *et al.* (2008, 2013), has been limited to focusing on known signals (e.g. the readiness potential) in known locations in the brain and has built on or simply recreated previous research. For the most part, research has not branched out to try find other and perhaps more reliable precursors of movement, nor has it ventured into other

regions of the brain that are not traditionally associated with movement preparation. This research stems from the unique findings of Soon *et al.* (2008, 2013) who identified markers for pre-conscious decisions in the frontopolar cortex. Therefore, this research will not undertake any assumptions as to where the "decisions" might arise, but rather investigate the brain as a whole. This will be achieved with the employment of machine learning algorithms.

In terms of analysis, most current methods are manual and require transformations of the data. Transformations include averaging which filters out much of the data to reveal "smooth" event-related potentials. Other techniques include statistical analysis to select pre-determined features such as the Root Mean Square [RMS] (Fergus *et al.*, 2015; Logesparan *et al.*, 2012; Volschenk, 2017); sample entropy, mean frequency (Fergus *et al.*, 2015; 2016) and signal entropy (Fergus *et al.*, 2015, 2016; Logesparan *et al.*, 2013, 2012). The limitation of the current methods of analysis as well as the focal EEG investigations (in terms of signals and location) are mainly due to the capacity limitations of manual data analysis techniques. Machine learning will not face this limitation as it is *almost* always scalable: in most instances it can work better as the amount of data increases (Géron, 2017). It can gain insights into complex problems especially those problems in which there is "no good solution at all using a traditional approach" (Géron, 2017), giving us an avenue for innovative research and discovery.

1.3 Aims and Objectives

1.3.1 Aim

The primary aim of this study was to investigate the distinction between the subconscious and the conscious brain and assign agency for our actions in order to determine the nature of free will.

The secondary aim of this research was to investigate the moment of conscious awareness and corresponding changes in the eyes.

1.3.2 Objectives

1.3.2.1 Primary objective

- To accurately classify a decision (left or right) by employing supervised learning neural networks with sub-conscious EEG data as the input

1.3.2.2 Secondary objectives

- To assess the degree to which the original experiment has been replicated by recreating the RP from the original Libet paradigm
- To determine the generalisability of the deep learning model when dealing with EEG data
- To assess the precision of a participant's subjective report of "W"
- To time-lock the moment of conscious awareness using eye-tracking

1.4 Rationale and significance of study

Research such as this contributes to the fundamental understanding of our brain's and our decision-making processes. Using the subconscious, the aim is to determine how early decisions are made. In other words, the investigation is essentially into the role of the subconscious in our decision-making and assigning agency to our actions and decisions.

This research has a potential impact far-reaching the age-old philosophical debate as to whether or not we have free will. Through the identification of new and earlier neural precursors (or "markers" for identifying movement) in other areas of the brain, this research could further lead to new developments in brain-computer interface (BCI) prosthetic-related research. If the BCI can understand (and in a sense predict) what the *brain* wants to do earlier on in the preparatory process, we can possibly improve the classification and execution of movement through BCI prosthetic systems. There is a current trend in neural engineering to develop more effective and efficient brain computer interfaces (BCI) with the specific intention of improving the lives of persons living with disabilities [PLWD] (Bai *et al.*, 2011). A BCI system uses the electrophysiological measures of brain function [viz. EEG, as fMRI has been found to be too limited in its "real-time connection capabilities" (Aggarwal *et al.*, 2008)]. These EEG signals are used to "enable communication between the brain and external devices such as computers" (Aggarwal *et al.*, 2008) or computer powered prosthetic/ exoskeleton type devices, and allows

the EEG signal to be converted into an artificial output (Nurse *et al*., 2015). The BCI provides a new/ different control pathway, should the normal physiological connection be severed as in the case of a spinal cord injury (SCI), for example. In this way, the PLWD can control the external device directly with their brain, as if it were their own limb.

The aim of BCI systems is to create movement that is as "natural" as possible (Bai *et al.*, 2011) i.e. to allow the PLWD to send out commands and have real-time movement or execution of an action; "while bypassing the brain's normal pathways of peripheral nerves and muscles" (Yang *et al.*, 2015) which may have been affected due to a brain injury (e.g. stroke), neurological degenerative disorders (e.g. multiple sclerosis [MS] and amyotrophic lateral sclerosis [ALS]), muscular degenerative diseases (such as Duchenne Muscular Dystrophy [DMD]), or even in the event of severe trauma resulting in permanent nerve damage or amputations (Aggarwal *et al.*, 2008) etc.

These systems are not without limitations, however, and there is much room for improvement in the prediction or interpretation of the person's intent (Bai *et al.*, 2011). Current BCI systems have focused on the motor cortex, under the assumption that this is the most logical starting point for neural signals to originate to produce a movement of sorts. This has meant that studies limit the scope of their EEG data acquisition to the areas directly over the motor cortex.

Specific difficulties include the prediction of what the person actually wants to do, in the event of considering whether to move or not over actually intending to definitely move (Bai *et al.*, 2011). This suggests the need for alternative, more accurate signals that can distinguish between consideration and intention to move thought processes. The EEG BCI systems are also limited to making one prediction every 100ms (Bai *et al.*, 2011), which results in latency in the detection of EEG based motor imagery (MI) (Ang *et al.*, 2015). This latency takes away the naturality of the movement, as early signals with the intent of movement can be missed. However, promising results have been found in studies using BCI to decode EEG signals earlier in the motor planning process (i.e. not just those in the motor cortex), highlighting the need for research to find earlier predictive signals for movement, in new unexplored areas of the brain.

2 Literature review

Chapter 2 summarises all relevant research into free will, electroencephalography (EEG) and eye tracking (both of which formed the basis of the data collection) and finally, machine learning and the branch thereof known as deep learning which was used as the main method of analysis for the EEG data.

2.1 Introduction to the literature review

In 2015, Neil deGrasse Tyson put forward that: "You have the illusion of free will… Because you are a prisoner of the present, forever locked in transition, between the past and the future". As mentioned in the previous section, our brains create our reality. Our "present" and thus "creation of reality" is not under our conscious control, but rather the result of subconscious neural preparatory processes. We don't know what these processes are doing nor what they are going to do. We are only aware of what they have decided to do and what they've decided to show us i.e. the forever "present".

Essentially, all this supports the notion that the preparation for movement begins long before our consciousness wakes up and claims ownership, agency or authorship of the decision to move. Therein lies the illusion – we are only aware of what our consciousness is aware of. David Eagleman sums this up perfectly: "Knowing yourself now requires the understanding that the 'conscious-you' occupies only a small room in the mansion of the brain, and that it has little control over the reality constructed for you". Research to date has identified regions in the pre-supplementary and supplementary motor areas (pre-SMA and SMA brain regions respectively) among others that are involved in these early stages of neural preparation, separating the roles of the subconscious and the conscious. However, it still has not been confirmed that these are the *earliest* precursory signals for movement.

This will be discussed in more detail in the coming paragraphs, along with theoretical explanations of the data collection tools and the data analysis methods.

2.2 Free will and the scientific exploration thereof

2.2.1 Products of process

As discussed, the seminal work of Libet *et al.*, (1983) paved the way for neuroscience to enter the debate on free will by providing empirical evidence to back up the theories of the philosophers, such as complete determinism, libertarianism and randomness. Most research has stemmed from the original Libet paradigm, in order to both support (and disprove) the idea that consciousness enters the process of preparation for movement after the brain has subconsciously decided what to do – i.e. we are the product of the subconscious workings of the brain, for which the conscious-self takes credit. The following sets out to expand on current literature, explaining the involvement of the subconscious on our own actions.

Disregarding the origin of a decision, to make an abstract choice, or to perform a movement, the conscious decision has certain defining traits (Krieghoff *et al.*, 2011). These differ from urges, or impulses in that specific "attention" is payed to a decision that is intended to meet a "desired goal", with an understanding of consequence (or "action-effect"). They are not automatic, and can be controlled (this will be discussed in detail in later sections) and require options in order to confirm with any certainty that the "preceding brain activity [doesn't] merely reflect the unspecific preparatory action" (Soon *et al.*, 2008). There are three components to a decision and any decision can encompass any number of these: "what (to do/ to decide upon), whether (or not to carry out the 'what') and when (should the conscious decision take form)". The Libet paradigm fails to meet these criteria and can be argued that the RP was preceding an urge. (Brass & Haggard, 2008)

The anatomy of a conscious decision is rather well understood, but it is the origin of these subconscious neural processes resulting in a decision that remain unclear. The origins of (sub)conscious decisions will be discussed next.

As mentioned previously, Soon *et al.* (2008) used fMRI to measure exactly which areas of the brain are involved in the early (subconscious) "shaping of a motor decision". A stream of letters was used to mark/ time-lock the moment of awareness of the decision to move ("W"). Participants could click with *either* their left or right index fingers. The time of movement was around $\pm$21.6 s after the start of the trial (one trial corresponds to one click with either finger). This meant there was enough time between trials to analyse the subconscious brain signals specifically and in isolation, as this time period between trials (i.e. 21.6 s) ensured that any

"predictive activity" wasn't remnant brain activity from the previous trial and the movement thereof. Researchers found "predictive [neural] information" in the pre-SMA and SMA seven seconds before "W" (as measured by the participant's report of which letter they saw at the moment they became aware of the decision to move); as well as in the frontopolar cortex and parietal cortex (precuneus in the posterior cingulate cortex) just before the onset of "W". Instead of using EEG, as in the original Libet experiment, the researchers made use of fMRI (functional magnetic resonance imaging). fMRI essentially looks at BOLD (blood-oxygenation level dependent) signals. This signal primarily responds to the changes in concentration of oxy – and de-oxyhaemoglobin [blood with and without oxygen](Huettel *et al*., 2004a). An increase in activity in one part of the brain causes an increase in oxygenated blood flow to this area – this leads to an increased ratio of oxyhaemoglobin to deoxyhaemoglobin. In this way, the fMRI is able to generate "a high-resolution image" of active and inactive neurons, as opposed to the measuring of the brain's electrical potentials as measured by EEG (Huettel *et al.*, 2004b). This process of capturing an fMRI image is somewhat sluggish (Aggarwal *et al.*, 2008), as the blood concentration ratio of oxy-and de-oxyhaemoglobin needs to change, allowing one to assume that these "predictive signals" were actually present up to 10 seconds before "W" (Soon *et al.*, 2008). There was also sufficient time between trials to rule out that these "predictive signals" were the lingering, remnant neural signals from previous trials. However, this and more recent studies are still unable to conclude if these were in fact the *earliest* predictors for movement, suggesting the opportunity to search for earlier neural precursors to movement (Fried *et al.*, 2011; Soon *et al.*, 2008).

This process of free decision making (being given options) was further explored along with the idea that similar results of predicting decisions could be applied to abstract decisions. An abstract decision is one that does not directly result in a physical movement, but rather remains a 'cognitive' process; i.e. an impalpable concept. Soon *et al*. (2013) conducted an experiment similar to the one discussed above. Instead of making a decision to click with either finger, the participants were asked to choose between two basic arithmetic operations – addition or subtraction. The stream of letters was used to time-lock the moment of the conscious awareness of the decision to perform the addition or subtraction. Besides aiming to see if prediction was possible, they were also looking for neural overlap (or lack thereof) between subconscious signals for decisions for movement and decisions for abstract intentions. The analysis of results focused on two determining points – the prediction of the timing of the decision as well as the prediction of addition or subtraction ("when versus what") decisions. The build-up of "when"

neural activity was found in the medial frontopolar cortex as well as the posterior cingulate cortex (precuneus specifically). The build-up of "what" neural activity was in the pre-SMA, the same area as the predictive neural activity for voluntary movement. In both cases, accurate predictions could be made as to whether the participant would perform addition or subtraction ("what") and "when" *before* the participant themselves were consciously aware of this decision.

A lot of current research into the question of free will look to old studies for methods to carry out their experiments. In doing so, research loses innovation and becomes stagnant in terms of simply "confirming" the results of previous studies (Johannesen *et al.*, 2016). Further, many features of EEG data are simply overlooked, out of ignorance and lack of novelty in analysing the data; and we are left with creative reproductions of prior work. However, the two studies above (Soon *et al.*, 2008, 2013) are set apart from the majority of other studies in this field, on account of their approach to the data collection and analysis. In both studies, the authors strayed from the traditional method of using electroencephalography (EEG) to look for the RP (Soon *et al.*, 2008, 2013). Instead, the authors, used fMRI and multivariate linear classification in order to classify actions up to seven seconds before conscious awareness. This innovation lead to a result better able to withstand the criticism of the Libet paradigm, as described in Chapter 1.

Fried *et al.* (2011) used the invasive technique of inserting depth electrodes into the pre-SMA, the SMA and the AAC. They initiated their experiment with the hypothesis that groups of neurons would work synchronously to bring about the decision to move (before this decision reaches conscious awareness). This study also found (besides concluding that the monitoring of a minimum of 256 neurons in the SMA is necessary to predict *subconscious* intent), that it doesn't matter that the participant's report of when "W" occurs is inaccurate. If the participant is within the margin of error of $\pm$ 200 ms (i.e. 200 ms off the actual moment of "W"), there is no significant change to the number of neurons "altering" subconscious brain activity preceding the estimated (and assumed) time of "W". In other words, even if the participant is inaccurate (within 200 ms) in retrospectively reporting the moment of "W", the neurons still engage in this subconscious activity, that relates to the decision to move before there is conscious involvement by the participants. The fault lies only in the measuring of the moment of "W", and not in what actually happens at this point, i.e. there is an exact moment of "W", and it doesn't matter too much if the participant doesn't recall it exactly. The researchers were able to predict which hand the participant would move, even before the participant themselves

were even aware of this decision. This prediction was possible regardless of how accurately the participant reported "W".

The studies of Fried *et al.* (2011) and Soon *et al.* (2008, 2013) assigned specific agency to the subconscious. Their conclusions are that the subconscious is in control of our decisions, and the conscious self merely witnesses the actions. Their conclusions suggest our conscious-self is not in control of the action, and that there is no free will. The following two studies also put forward that there is no free will, but in contrast, do not assign agency to the subconscious. Murakami *et al.* (2014) and Schurger *et al.* (2012) suggest all actions are the result of random neural fluctuations crossing a threshold. In other words, they propose the subconscious is not acting independently, it is subject to random neural activity.

Schurger *et al*. (2012) introduced a new model of subconscious neural processing called "stochastic accumulation". Their motivation behind developing a new model is simple: all studies centered around the RP focus only on the "last one to two seconds before movement onset", but almost all disregard subconscious brain activity when there is no specific movement or decision put into question. The "stochastic accumulation" refers to the random fluctuations of neural signals that either move toward or away from a "threshold". This threshold is synonymous with the crossing from subconsciousness to consciousness; i.e. "W" (Murakami *et al*., 2014). In order for movement to proceed or for a subconscious intent to manifest itself, these randomly fluctuating signals need to rise above the threshold. These fluctuations are inherently spontaneous and random, and there is no way to control it. These signals may come very close to the threshold, but not cross it – this is a purely random occurrence, there is no way to determine nor influence whether or not the signals will cross the threshold.

In order to come up with their model, Schurger *et al.* analysed reaction times to a cued movement in order to come up with their "stochastic accumulator model". Their conclusion was that if a cued movement occurs quicker than another similar cued movement, then the spontaneous neural signals were fluctuating close to the threshold already, and thus the decision to move could occur faster than if the threshold first had to be reached from lower fluctuating signals. The crossing of the threshold is synonymous to the awareness of the decision to move entering consciousness; i.e. crossing the threshold corresponds to "W".

The aim of Murakami, *et al.* (2014) was to investigate the underlying cause for these random, neural fluctuations. Where do the signals for spontaneous decisions originate in the absence of an "external trigger"; i.e. a purely self-initiated decision? Murakami, *et al.* (2014) used rats

to conduct their study, focusing on the self-initiating task of "deciding when to give up waiting for an anticipated event whose timing is uncertain". This involved a waiting task that ended in a reward signalled by a tone. The rat could go for the reward (water) after the first tone (which occurred at a fixed time point), or wait for a larger reward, but the timing of the second reward was uncertain. Electrodes were placed on the rostral secondary motor cortex (M2) to assess the activity of neurons. Individual neurons that showed "ramp to threshold activity" were those that were pushing for the abortion of waiting time and settling for the smaller reward. The neurons that didn't show this, had different firing rates and were pushing to wait for the larger reward. Each individual neuron showed its own firing rate, but whether or not the rat would wait for the larger reward depended on the number of neurons pushing for a certain decision. Neurons would "[co-ordinate] until their combined activity crosses a threshold". Essentially, it becomes a tug of war between neurons - the decision that ultimately occurs (abort or wait) is subject to which group of neurons over-powered the others. The integration to bound model incorporates the randomness of this as well as show strong links to a possible explanation for impulsivity.

Sam Harris (2012) puts these random fluctuations into context for us – consider a case of an attempt at impulse control such as dieting or trying to quit smoking. Many people embark on either of these endeavours, and some attempts are more successful than others. What exactly separates the case of actually sticking to the diet plan that you start this year, or finally managing to stave off the nicotine cravings, as opposed to trying and failing the previous three years? The answer is simple: random chance. If we are indeed the product of random subconscious neural fluctuations, we have no control over this. It is merely a case of chance that this is the time you are able to stave off doughnuts, exercise every week or finally kick the smoking habit. This is the time that the neurons pushing for healthier lifestyle surpass the power of the neurons pushing for the couch or cigarette, and it is the result of the firing rates of the pro-health neurons that result in the (still random) spontaneous fluctuations crossing the threshold to result in the decision to reach for the celery instead. Unfortunately (or fortunately in the right case), it remains completely random and uncontrollable by you as to which neuron group will win.

Assigning agency to chance, to the independent self, or to the independent subconscious is important in the debate of free will. Understanding the cause of our decisions can help us better understand the factors that govern our actions. This idea, of being a product of processes (i.e.

the resultant product of subconscious brain activity) is discussed in different applications and settings in the next section.

2.2.2 The game of chance

With the exception of Murakami *et al.* (2014), the studies described above didn't have any elements of consequence. In the experiment of Murakami *et al.* (2014), the rats' decision lead to a reward. i.e. there was a consequential event. The decisions investigated in the experiments of Schurger *et al.* (2012) and Soon *et al.* (2008, 2013) are still far removed from real-world scenarios.

The work of Bechara *et al*. (1997) and Maoz *et al.* (2017) sought to investigate decision-making processes in the context of consequence. In 1997, Bechara *et al*., conducted an experiment in which participants were given four decks of cards and $2000. The only thing explained to them was to try and win as many and lose as few points as possible. Turning a card would either result in a reward or loss. Unknown to the participants, choosing from decks A and B were overall disadvantageous – along with the high reward came high penalties. Decks C and D were overall advantageous, with lower rewards and lower penalties. Skin conductance responses (SCR's) were measured and patients were asked after the first 20 moves (and then every 10 after that) about their understanding of the game and the strategy they were employing. What the researchers found was that even before the participants had a full understanding of the game, and which decks were advantageous, they began to "generate" SCR's before drawing from the disadvantageous decks A and B and subconsciously began avoiding these as well (despite reporting in the intermittent interviews they didn't have a full "conceptualization" of the game). Thus, what these results suggest is that the subconscious had picked up the understanding of which cards would lead to better results in the long term, even before the "consciousness" had realised this.

Maoz *et al.* (2017) set out to compare "neural precursors of deliberate and arbitrary decisions" through the analysis of EEG signals. Their motivation for comparing "deliberate and arbitrary decisions" comes from the point raised right at the beginning: in order for a free choice to be distinguished from a reaction, or urge, there needs to be some consequence related to it. The researchers recruited 18 participants who were active, knowledgeable members of society with a history of donating to charity and voting in elections. The participants were presented with two non-profit organisations (NPOs) at a time. They could choose which NPO they would like to donate $500 to 50-real NPOs were presented: 20 were similar in the sense they represented

cancer research and hunger programmes and the other 30 were more "controversial: … widely debated topics such as pro-/anti-abortion or pro-/anti-gun laws". NPO pairs selected for this category represented each side (pro- / anti-) of each debate. In "deliberate decision" trials, one NPO would be selected over the other to receive the money, but in the "arbitrary decision" trials, both NPOs would receive the same amount of money, regardless of the choice. The participants were under the impression these organisations would receive the money, and in this way "real-world consequence" was introduced into the experiment. Their results showed the RP to only be present in arbitrary decision trials. Further there were different origins of neural precursors for deliberate (prefrontal cortex, AAC), and arbitrary trials (SMA, posterior cingulate cortex). These results clearly show the flaw in trying to generalise the signals that precede arbitrary decisions (i.e. lifting either hand) to signals that precede real world consequential decisions; this also reinforces the notion that different subconscious processes govern different kinds of decisions and emphasises the need for consequence in the development of a strategy to try elicit the correct neural signals for volitional decisions. With the little understanding of the subconscious that we have, we are still able to marvel at its capabilities. The following study presents the power of the subconscious in a situation that meets the criteria set out by Maoz *et al.* (2017): a real-life scenario with real-world consequence – gambling.

This result of the RP only being present in 'arbitrary" decisions is further evidence of the need to move away from the RP analysis in the context of free will and decision-making related research. Further research is however needed to understand the factors influencing our actions.

2.2.3 Offsetting the subconscious

It is difficult to accept that the subconscious is the only force governing our actions. Are there any conscious forces that are able to govern the will, or process, of the subconscious? The following explores these possibilities:

Benjamin Libet commented on his own seminal work in 1999, suggesting that the consciousness can have a "veto" or over-riding effect on the actions or decisions starting in the subconscious. Libet proposes that there is a play between the conscious and subconscious mind - the preparation for movement begins subconsciously, while the conscious part of the brain decides whether or not this action can take place. Subconscious intentions "bubble up" to the point of consciousness awareness, and the consciousness "elects which of these may go forward". However, this doesn't completely exclude the influence of the subconscious and

definitively prove we are free-willed beings, as we have no evidence proving that this "conscious veto" appears without its own development of an "unconscious origin". (Libet, 1999) This "vetoing" of subconscious neural events is in line with the idea of compatibilism, in that some actions are free and some are the product of subconscious brain activity. This "veto-power" will be discussed in more detail next.

Schultze-Kraft *et al.*, (2016) had participants play a game against a BCI system. In this game, participants had to press a button with their right foot whenever they wanted. They would receive a point if they pressed while the light was green, and lose a point if they pressed when the light was red. Initially, the light would turn from green to red randomly. In the second round, unknown to the participants, the BCI system began analysing their EEG signals to try and predict exactly when the participant would press with their foot. In this way, the BCI could predict when the command would go from the brain to the foot to move, and then turn the light red at the last possible moment so that it would be too late for the participant to stop the movement. In that event, the participant would lose (having pressed while the button was red), and effectively the BCI system won. In the final round, the participants were made aware of what the BCI system was doing and instructed to try behave "unpredictably". In the third round, it became a game of trying to "veto" a ready-made decision to move and then not move. The BCI was able to monitor the subconscious brain signals and predict when the participant would move, before the participant was aware of this. However, the idea was for the participant to try trick the BCI system; they would become consciously aware of the decision to move ("W") and then not execute the movement at random instances. The study found that the participants were able to "veto" 200 ms before the onset of movement ("M") and no activity was recorded on the electromyography (EMG); concluding "it is possible to change or abort one's movement" after the onset of the RP. A veto implies no physical reaction whatsoever. Later than 200 ms, the initial subconscious decision isn't aborted, but rather changed and the EMG will still pick up some kind of activity. This is known as "late cancellation".

The above study shows there is potential to offset the subconscious, but as Libet, (1999) pointed out – there is no way to know if this conscious feeling of vetoing the products of subconscious brain activity is in fact a product of the free-willed "self", or the separate subconscious controlled brain activity. This brings us back to the original question – what possesses the final control over our actions?

2.2.4 Blurry rose-coloured glasses

The original Libet paradigm and subsequent studies making use thereof have been criticised regarding the validity of the conclusions – the report of "W" is too reliant on the participant's ability to focus on the task at hand, while also remembering a moment in time that they became consciously aware of this decision to move (Jo *et al*., 2015). The report of "W" is also given retrospectively, once the trial is completed; hindering the preciseness of this retrospective, subjective report of a moment in time. The following studies assessed this retrospective nature of reporting "W" and determine whether or not the perception of when "W" occurs can be altered through external cues and stimuli.

Lau *et al*., (2007) first introduced the idea of "manipulating the experienced onset of intention" after the action had been completed. Since all reports of "W" occur after the action has been completed, the participant's perception and recall play a role in the report of this moment in time. Their aim was to see how much the participant's perception of time could be altered by external influences. In this case, direct trans-cranial magnetic stimulation (TMS) was delivered to the pre-SMA at a random time during the standard Libet paradigm procedure. The study found that the "perceived onset of intention can be manipulated by the TMS as late as 200 ms after the execution of a voluntary *action*" (i.e. a stimulus delivered 200 ms after "M" can influence the person's perception of the moment in time that "W" [awareness to intention] occurred).

Banks & Isham (2008) were also able to manipulate their participants reporting of "W". By introducing timeous tones or delayed video feed, the authors were able to manipulate the participants' perception of time. The participants, on occasion, recorded "W" to occur after "M". This confirms the inaccuracy of reporting the moment of conscious awareness after the trial has finished – as it is vulnerable to be manipulated by our environment and memories.

As shown above, the retrospective report of "W" can be flawed, for a variety of reasons. It is for this reason, that an innovative approach to measuring the moment of conscious awareness is necessary. This will be discussed in detail later. Despite efforts to address these limitations in precisely determining "W", there are those that claim their research has completely disproved Libet *et al.* (1983). This will be discussed in the following section.

2.2.5 Dissention

Trevena & Miller (2010) sought further insight into the readiness potential, suggesting that perhaps the RP was not specific to movement preparation, but rather it is an indicator of "on-going [background] brain processes" where attention specifically is employed. Therefore, the RP is not specifically "event-related", but a rather vague signal and the claims of Libet et al. (1983) cannot hold true. In an attempt to remove the subjective report of "W", the researchers dictated this moment with a tone. The first group of participants had to decide, at the tone, whether they would move or not (decision-only trials). The second group of participants had to wait until the tone to decide which hand to move, and then move it (decision and action trials). The first group merely made a decision, while the second carried out their decision (i.e. a decision followed by action). Their conclusion was that they found no difference in the size of the RP signal in the EEG – comparing the decision, versus the action. They conclude that the RP is not "event-related" on account of it being the same in both instances. However, these researchers adapted the Libet paradigm quite drastically. These adaptations could in themselves be explanation for the different results, and their results are also countered by the findings of Soon et al. (2013) which specifically focused on abstract intentions.

Miller *et al.* (2011) found the RP to be a mere artefact of monitoring the Libet clock. An artefact on EEG is a signal that arises for reasons other than pure cerebral activity - such as a physiological heartbeat or blink, or in this case a non-physiological extraneous source: the monitoring of the clock used in all Libet paradigm-styled experiments (Schomer & Lopes da Silva, 2011). This experiment followed the Libet paradigm, where participants were first asked to follow the standard procedure: monitor the clock to time-lock, "W", perform action "M" and then report the time of "W". The participants were then asked to only focus on "M" – i.e. no monitoring the clock, nor was any attention to be paid to "W". Researchers only found the "negative potential", viz. the RP, in the trials where the complete Libet paradigm was executed. Their conclusion presents the opportunity for further research: current research relies on the assumption that the RP is specific to movement preparation, but their research has found a flaw in this. In order to definitively separate the neural signals for "clock monitoring and movement preparation", EEG recording and analysis needs to become more "refined". However, one could argue that refining the process will still be limited unless a larger scope of the brain is explored, with the intent of finding new neural precursors to movement.

2.2.6 Addressing the criticism

To date, there have been experiments focused on improving various aspects of the Libet paradigm. One of the challenges is accurately separating the subconscious from the conscious – objectively time-locking the moment of conscious awareness of a decision ("W"). The original Libet paradigm involves the participant subjectively reporting this "W" after the trial has finished. The subjective reporting of "W" has already been proven to be vulnerable to manipulation (Banks & Isham, 2008; Lau *et al.*, 2007), highlighting the need for a more accurate measure.

Jo *et al.*, (2015) and Jo *et al.*, (2014) found meditators to have greater attention and greater "capacity for attention to their inner purposes" as compared to non-meditators. Meditation is a mindful process and involves "focusing one's mind for a period of time for spiritual purposes or as a method of relaxation" (Oxford South African Concise Dictionary, 2010, s.v. 'meditate'.). This finding is advantageous, as it proposes a more accurate means of reporting "W". However, their studies focused mainly on the point of "W" and the inner processes around that point in time and did not focus on subconscious signals preceding this moment of awareness. Further, "experienced" meditators practise with dedication for many years, and using this as an inclusion measure would eliminate the majority of people. This is a very niche inclusion criterion – making the research inapplicable and non-generalisable to the population as a whole. This, however, is not a suitable solution to the problem – the report will still be inherently subjective (and those always vulnerable to scrutiny) and not a reliable solution to the problem.

Evidence shows that eyes correspond to cognitive processes (Anantrasirichai *et al.*, 2016; Blignaut, 2009; Salvucci & Goldberg, 2000). Gaze fixations are of especial interest. During a fixation, a significant amount of cognitive processing takes place – and while we may not be *consciously* aware of all the information our brains take in during these fixations, these fixations provide a convenient window into our visual and cognitive processes. These fixations can be timestamped and give us an indication of "when" cognition was stimulated by "what" (Anantrasirichai *et al.*, 2016).

Eye tracking provides a promising avenue through which the development of more objective measures of time-locking "W" can be investigated. Eye tracking is described in detail in section 2.4.

2.2.7 Summary of the free will component

As evidenced above, there are those who have claimed to both prove and disprove that which initially Libet proposed – that the preparatory processes for movement begin before conscious involvement plays a role and therefore there cannot be free will. However, and as noted by others, these studies are still limited in various ways.

It is necessary to develop an experiment that is free of subjective involvement, which relies purely on objective measurements of the person's attention to awareness. This experiment that looks at all parts of the brain, not just those already identified. Without these key factors it will be difficult to conclude that the brain acts (or does *not* act) independently to the conscious self.

This is exactly what this research intends to do through the combined use of EEG, machine learning and eye tracking. Each of these tools will be described in detail in the proceeding sections.

2.3 Electroencephalography

2.3.1 What is electroencephalography (EEG)?

Electroencephalography (EEG) is the non-invasive recording of the brain's electrical activity from the scalp. This electrical activity is the result of ionic flow between neurons (via neuro-transmitters) through the extracellular space which creates a change in electrical potential, which in turn creates an electrical field that surrounds a neuron. The EEG is not able to record the electrical signals of a single neuron's activity, but, rather requires the synchronous excitation and summation of hundreds/ thousands or tens of thousands of neurons. (Cohen, 2014)

It must be noted at this point, that these neuronal electrical fields obey physical laws (Schomer & Lopes da Silva, 2011), and are not under our own conscious control.

In order for EEG to record local field potentials it requires the simultaneous firing of an ensemble of neurons (EEG cannot record signal from neuronal ensembles with spatial areas of <3 mm^3 [i.e. in the microscopic scale]). Patches of cortex in the mesoscopic scale (i.e. >3 mm^3 to a few cm^3) can be measured by a high density of at least 64 electrodes. EEG records most easily from groups of neurons in macroscopic scale - > a few cm^3, with less than 64 electrodes, however, if the aim is to make deductions about brain localisation one requires at least 100

electrodes (Cohen, 2014). EEG struggles to record field potentials from deep brain structures as field strength decreases as an exponential function of distance (Cohen, 2014).

As a general rule, signals that arise from a random orientation of neurons will not be recorded, regardless of depth, even assuming a large synchronous origin in terms of neurons firing. Furthermore, the EEG will only record signals from neurons with parallel terminal ends (which extend from the axon of the neuron, as depicted in Figure 3 below). It is also required that these terminal ends are perpendicular to the surface of the scalp (Schomer & Lopes da Silva, 2011). Figure 3 depicts a typical neural and Table 1 provides the key.

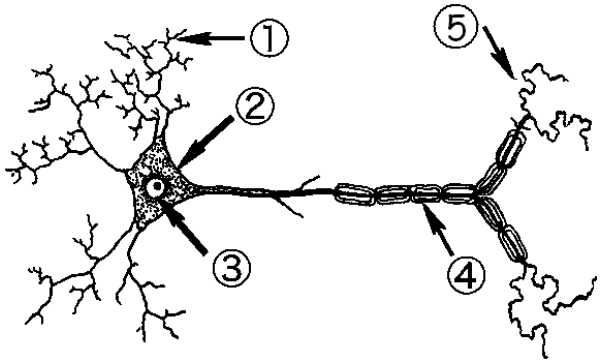


Figure 3: Diagram of a typical neuron (Schmidt, n.d.)

Table 1: Legend for Figure 3

Key	Explanation
1	Dendritic trees
2	Neuronal cell body
3	Nucleus
4	Axon
5	Terminal end: dendritic trees

2.3.2 What makes up an EEG signal?

Oscillations are waves occurring in the EEG that are a result of fluctuations in the excitation of groups of neurons. The oscillation is the result of shifts in electrical fields (which surround neurons) between the states of excitation and inhibition. The fluctuations in the state of the electrical fields is due to interactions between the central nervous system's (CNS) primary excitatory units [pyramidal cells] and the primary inhibitory neurotransmitter [gamma (γ)-Aminobutyric acid (GABA-Aergic)]. As mentioned, these oscillating electrical fields are too weak individually to be measured by the EEG system, but when summed across an entire

"population" or group of neurons, these signals become strong enough to travel through brain tissue, the blood-brain barrier (semi-permeable membrane separating blood and the cerebrospinal fluid(CSF)), skull, skin and hair to be recorded by the electrodes placed on the participant's head. (Cohen, 2014)

These oscillations are described in terms of:

- Phase: position along the sine wave at any given point, measured in radians/ degrees
- Power: amount of energy in a frequency band, measured by the squared amplitude of the oscillation
- Frequency: the speed of the oscillation – i.e. the number of cycles per second, measured in Hz

Figure 4 provides a graphic representation of the power and frequency characteristic of EEG.

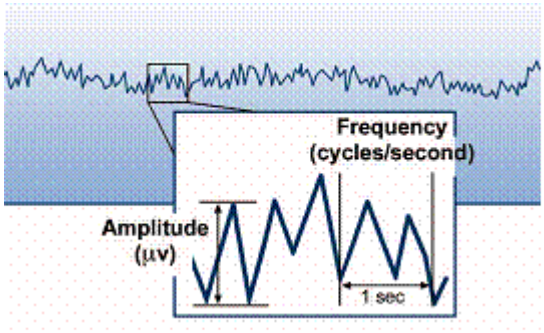


Figure 4: Graphic representation of the power ("amplitude") and frequency (Arnesano, 2009)

Table 2 provides the typical EEG frequencies of different brain waves. These values are taken from Cohen, 2014. There is little convention designating where one signal starts, and another ends (Volschenk, 2017): but, the ranges are all comparable.

Table 2: Frequency ranges and graphic depictions of brain activity

Wave	Frequency
Gamma (γ)	$30 - 100 +$ Hz
Beta (β)	$14 - 30$ Hz
Alpha (α)	$8 - 13.9$ Hz
Theta (ϕ)	$4 - 7.9$ Hz
Delta (δ)	$0.1 - 3.9$ Hz
Brain death	0 Hz

Values taken from Cohen (2014)

2.3.3 How are these signals measured?

As measured, the brain tissue (although not in isolation) is able to produce electrical activity. Volume conduction refers to the movement (conduction) of this electrical activity (EEG in this case) through the body's tissue and bones to be measured by electrodes, placed over a person's scalp (Holsheimer & Feenstra, 1977). Each electrode is most sensitive to the electrical activity directly below it (as well as other sources such as electromyography [EMG]), however, this sensitivity can be negated by certain factors: air has zero conductivity, while the scalp's cutaneous oils, keratin and epidermis are good insulators. Other factors influencing the strength of EEG recordings include: sweating on the scalp and/ or movement of the electrodes. It is for these reasons that the electrodes are placed as close to the scalp as possible. Different EEG systems have different measures in place to mitigate the effects of the factors listed above, and to improve the current flow from brain, through the scalp and hair to the electrodes. This study will make use of the BRAIN Products ActiCAP EEG system with 128 gel electrodes. The gel is a kind of electrolyte (Na^+ and Cl^- [sodium and chloride]) that improves conductance of neural electrical activity to be measured by the EEG system. (Schomer & Lopes da Silva, 2011)

Electrodes are placed on the scalp according to the international 10-20 system, developed by Dr H.H Jasper in 1958. The image (Figure 5) below depict standardised 10-20 system. Table 3 provides the key for Figure 5:

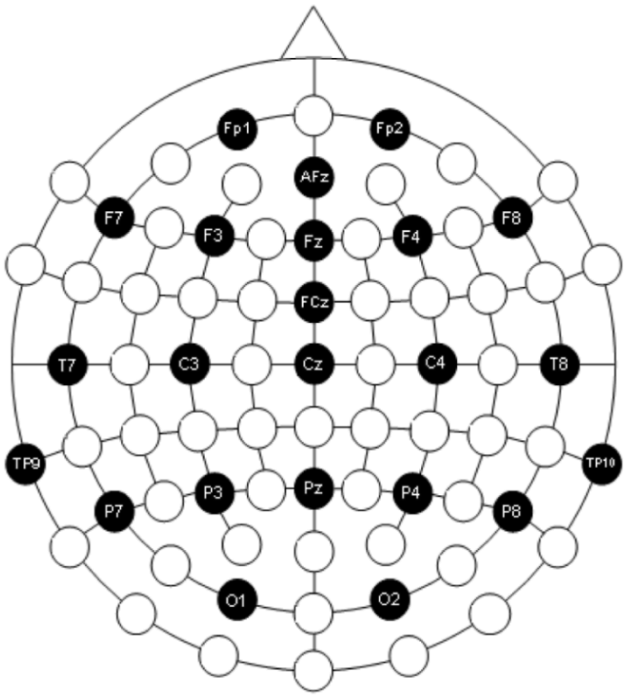


Figure 5: International 10-20 system for electrode placement with nomenclature ("Template 2D layouts for plotting", 2018).

Table 3: Legend for Figure 5

Key	Explanation
First letter	Brain lobe
C	Central
Fp	Fronto-polar
F	Frontal
T	Temporal
P	Parietal
Subscript (number)	Indicates the side of head: Odd = left, even = right, Z = midline. The more lateral you go, the higher the number.

The 10-20 system is so named after the percentages refer to the distance allocated to different electrodes between the main bony landmarks, namely the nasion (nasal point between the eyes) and the inion (bony prominence above the occiput) as well the bilateral pre-auricular (anterior to the ear) points.

Each electrode can pick up overlapping signals – i.e. more than one electrode can pick up the same brain activity as another, contiguous electrode. However, the strength of the signal is inversely proportional to the distance from the signal, so the electrode closest to the source will pick up the larger signal (Cohen, 2014). This phenomenon is mitigated through the use of amplifiers. Amplifiers, otherwise referred to as "differential amplifiers", measure the voltage difference between two signals at each of its inputs (i.e. electrical activity inputted to two separate electrodes).

The formula used by EEG amplifiers is as follows:

$$EEG\ signal = (input\ 1) - (input\ 2)$$

Equation 2.1

(Schomer & Lopes da Silva, 2011)

Essentially, what this means is that two electrodes each record a separate signal, however, there may be overlapping sources of signal in each. The two electrodes pass their respective signals through the amplifier and the result is outputted as a signal EEG channel reading (EEG signal in the equation above), which will be seen on the computer monitor, as depicted below in figure Figure 6:

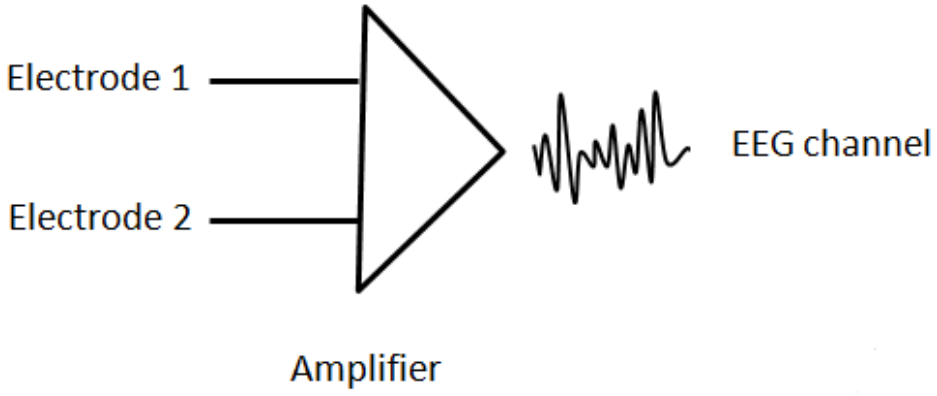


Figure 6: The process of signals being passed through an amplifier and outputted as a single EEG channel (Smith, n.d.; Zakeri, 2016)

The advantage of this, is that if two electrodes do measure the same signal, it won't be represented twice – these identical signals will be cancelled out. This is also particularly useful for the cancellation of external noise measured by contiguous electrodes.

The resultant voltage (measured in microvolts (μV)) is displayed as the size of the distance the signal deviates from the zero baseline (i.e. the amplitude – an example of which is depicted below in Figure 7).

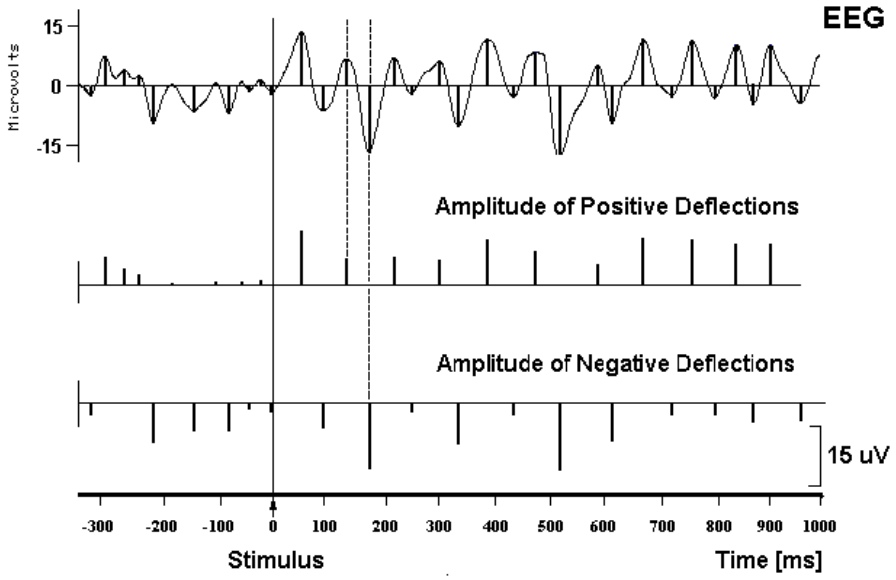


Figure 7: Diagram illustrating positive and negative deflections of various amplitudes from a zero baseline (Goodman, 2002)

2.3.4 Does the EEG only measure cognitive activity?

EEG, as it is, is susceptible to recording noise (Cohen, 2014). Noise is created, and picked by the EEG from sources other than the brain, creating what is also known as artefacts. Artefacts sum on top of the cognitive data.

These artefacts can be physiological in origin (i.e. arising from a source in the body), or non-physiological (i.e. an external source). Examples are given as follows:

- Physiological:

  Eye movements (including blinking), glossokinetic (tongue movements), electrocardiographic (from the heart), sweating or patient's movements (clenching their jaw, fidgeting). (Cohen, 2014)
- Non-physiological:

  Instrumental (electrode placement, amplifiers, environmental, brief amplifier saturations and line noise at 50 or 60 Hz. (Cohen, 2014)

There are several simple steps one can take, to pre-empt or, at least, minimise artefacts. These measures include instructing participants to not use gel or conditioner in the hair the night before/ day of the study and ensuring comfort (in terms of room temperature, seating position etc.) during the trial to minimise fidgeting and sweating. Further, demonstrations of how EEG can be contaminated by simple acts such as clenching their jaws, or giggling may help the participant understand the importance of avoiding these during EEG recording, reducing potential sources of artefacts and thus reducing the need for extensive pre-processing.

EEG forms the basis of data collection in the Libet paradigm in order to better understand the role of the subconscious in decision making processes. Eye tracking is also important in the context of this research and will be described in the proceeding section.

2.4 Eye tracking

Eyes are a means of accessing our attention mechanisms, including "visual attention" (Duchowski, 2007; Tobii Pro, n.d.). Theories around visual attention have been documented for almost 100 years. The following quote by American philosopher and psychologist William James (1981) is useful in understanding what is meant by visual attention:

> *Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalisation [and] concentration of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others...When the things are apprehended by the senses, the number of them that can be attended to at once is small,* 'Pluribus intentus, minor est ad singular sensus.'
>
> James, 1981

The Latin above roughly translates to "*Many filtered into few for perception*." (Duchowski, 2007). Essentially, what this all comes down to is when viewing the world, so much is taken in, but we can only be *consciously* aware of so much. That is, our eyes take in the visual world, passing information along to the brain and depending on the demand of the entity of interest, this attention can be more focused or broader. Essentially that which we perceive or consciously register, is only a selection of our filtered-out version of the world.

"Visual attention" refers to the (visual) signals that manage to reach our conscious awareness from the eyes having passed the brain's cognitive-processing "selection/ filtration" centres.

This develops the way we experience our world (the conscious experience). This idea of "visual attention" also clearly separates our subconscious, which takes in all the information and filters it, and the conscious which is our reality

Eye tracking (ET) is a non-invasive form of studying "visual attention" (Duchowski, 2007; iMotions, 2018; Tobii Pro, n.d.). As already mentioned, the eye reacts to changes in cognition and attention, be it change in where we look (focus our gaze), or responsive pupil constriction or dilation in response to an attentional stimulus (Einhäuser *et al.*, 2010; Wierda *et al.*, 2012) . Eye tracking continuously records movements and thus capturing both conscious and sub-conscious eye movements. In other words, eye tracking records all eye movements; tracking where we look, even though we don't consciously *see* anything per se.

A single limitation, that must be noted, is that the eye tracker in isolation cannot capture what we perceive (Duchowski, 2007; iMotions, 2018) or what unconscious brain process is causing us to look where we do. It merely catches eye movements. Beyond that, inferences would have to be made by understanding the relationship between decisions and the corresponding changes in the eyes.

To understand this notion of seeing, but not perceiving, consider looking for a pen. The pen is right in front of you, it's been there all the time. Your eyes passed over it, printed the image onto your brain, but, consciously, you are frustratingly unaware it is right in front of you. It's there, you've seen it, but maybe your subconscious has decided to play a trick on you and not let you in on the secret of the pen's whereabouts. Eventually, having given up you get a new pen from your drawer, only to look back at your desk to find your pen staring you down. All the while your subconscious mind sits back and chuckles at its cunning.

In terms of the research at hand – where the reporting of the moment of conscious awareness ("W") of the decision to act is a retrospective subjective report. There is criticism in the validity of the methods aimed at time-locking "W". In order to avoid this, this research proposes that the eyes may reveal an indication of awareness when looking at the clock – such as a hypothetical increased duration of the eye's focus at a point in time. Eye tracking will be used to investigate its role in objectively marking the moment of conscious awareness.

To aid the understanding of visual attention, a break-down of the anatomy of the eye, as well as a complete description of eye movements can be found in the appendix; in sections A.1 and A.2, respectively. The criticism of the use of the RP as conclusive evidence of the role of the

subconscious in our decision-making processes has come under a lot of criticism. The use of machine learning in this research is to address the criticism.

2.4.1 How eye tracking works

The basis of eye tracking systems is to record reflections off the surface of the eye. The first recordings of measuring "corneal reflections" are as early as 1901, with the use of contact lenses as a medium for reflections being introduced in the 1950s. Mirrors and/ or metal coils were used to measure reflections. Although the direct contact of the contact lenses improved the sensitivity of the devices, it was invasive. Modern systems are non-invasive and are termed "remote" eye tracking devices.

Eye tracking systems are centered around PCCR – pupil centre/ corneal reflections (Duchowski, 2007). A light source, typically infrared (Duchowski, 2007; iMotions, 2018; Tobii Pro, n.d.) is directed to the eye causing a reflection on the cornea. This reflection is continuously measured relative to the position of the pupil centre. Two points of reference are needed to separate eye movements from head movements - head movements would result in uniform shift of the cornea and pupil. A movement of the eye would result in a change in the "positional difference" between the pupil centre and corneal reflections. (Duchowski, 2007) These points are continuously recorded by a camera, resulting in the real-time update of the eye's position, by calculating a vector of the angle between the two points of interest.

Remote eye tracking can be in the form of (portable) glasses or a screen-based bar. The screen-based bar is best used for experiments involving stimuli from the computer monitor (iMotions, 2018; Tobii Pro, n.d.); such as will be present in this research project with the display of the Libet clock being displayed on the computer monitor.

This research will be making use of Tobii eye tracking hardware which uses a similar principle to PCCR, with improved technology. The light source used is near-infrared with image sensors capturing images of the eyes and reflection patterns. The software then uses this information and inputs it into "advanced image processing algorithms and a 3-D model of the eye" to estimate the position of the eye in space and the point of gaze with high accuracy. (Tobii Pro, n.d.)

Research has shown that pupil dilation and fixations correlates to attention (Anantrasirichai *et al*., 2016; Blignaut, 2009; Wierda et al., 2012). The algorithms in the Tobii ET device can also be used to calculate a relative change in pupil size; as well as the position of each pupil based

on an internal co-ordinate system. These measurements are however not as accurate as that measured by a pupillometer. It is important to note that the Tobii devices can be used to monitor "variations in size over time" but not the actual size, as the output of the algorithms is a "model generated value in mms or arbitrary units". Therefore, all participants will be measured with the same device in order to produce reliable measures. (Tobii Pro, 2015)

When recording eye tracking data, it is possible to introduce noise into the data, as with EEG. Sources of noise are generally other reflections caused by light in the visible spectrum – such as bright lights or direct sunlight. Spectacles can cause alternative reflections recorded by the ET device, and so removal thereof decreases the potential for noise.

As seen above, introducing eye tracking into an experiment typically reserved for brain imaging techniques is a useful means of gain insights into the underlying brain processes we are not consciously aware of and has the potential to improve the measuring of the moment of conscious awareness ("W"). The next component used to investigate the role of the subconscious in decision-making is machine learning.

2.5 Machine (and deep) learning

Machine learning, and the branch thereof known as deep learning, is presented as a solution to the criticism the original Libet paradigm has faced in terms of EEG data analysis. However, before addressing these criticisms, the following explanation of machine learning is introduced.

Arthur Samuels, a pioneer in the field of artificial intelligence, first introduced the term "machine learning", in 1959, defining it as: "giving the computer the ability to learn without being explicitly programmed"(Samuel, 1959). In 1998, Tom Mitchell, the Professor of machine learning at Carnegie Mellon University, expanded on this idea by bringing the following explanation forward:

> *A computer is said to learn from experience E, with respect to some task T, and some performance measure P, if its performance on T, as measured by P, improves with experience E.*
>
> Mitchell, 1998

Essentially, machine learning gives the computer the 'ability' to learn patterns from the data and output this as useful information for the programmer. Deep learning gives the computer a greater ability to learn more complex relationships between the data.

Deep learning, or the formation of neural networks, is a type of machine learning in which there are multiple layers. The capacity for increased problem solving comes from the addition of layers (of functions) to the framework. These "hidden layers" (between the input and output layers) allow for a greater potential of problem solving, as each layer processes the information fed into it slightly differently, resulting in the learning of complex representations of the information. (Goodfellow *et al*., 2016)

There are various kinds of machine learning, each advantageous depending on the application. As established, an innovative means of analysing data need to be explored and utilized in revolutionising the neuroscientific approach to free will. That is, to develop more valid methods in investigating the role of the subconscious in our decision-making processes.

The analysis of EEG data through machine learning will assist in this process. It is for this reason that machine learning (or, more specifically, the branch thereof known as deep learning) will form an integral part of his study.

2.5.1 What are the specific advantages of machine learning?

Machine (and deep) learning generally supersedes human capabilities, as well as traditional hard-coded methods of data analysis, in that it learns from the data and isn't always bound to human logic.

Machine learning is especially useful in solving tasks that are either too complex for "traditional approaches", or when it is not known which is the most apt model for a task, or if a suitable machine learning model even exists (Géron, 2017). For clarification, model can be used interchangeably with the term 'algorithm' in the context of machine learning.

Another motivation for using a ML system is that they can handle "fluctuating environments" (Géron, 2017), a key requirement in the context of EEG data analysis, in that it cannot be guaranteed that everyone's EEG patterns will be the same, nor can it be assumed that each trial for one participant will be identical, due to confounding factors such as fatigue, hormonal changes and/or concentration levels or inherent noise (of varying origin) in the EEG data.

This potential for innovation is essential in the context of this research. Neuroscientific research looking into the question of free will has maintained a consistent approach since the seminal work of Libet *et al*. (1983). Machine learning doesn't follow "hard-coded logic" (Géron, 2017), which can be subject to 'human bias'. Rather, ML looks at the data for what it is, free of any

pre-conceived ideas. This lack of bias (pre-conception in this context) allows the model to find what "rules" , or patterns or features it can (Géron, 2017).

Another criticism of the original Libet paradigm is the approach to EEG analysis is the averaging of the data across many trials. The averaging smooths out the signals to reveal ERPs (event-related potentials). However, in this process, we "collapse the dynamic information" in the EEG – essentially taking out the unique differences in different trials, and limiting the potential to make inferences about the brain processing as we have only kept that which is common (average) among all trials (Delorme & Makerig, n.d.). In other words, the RP is not present on a single trial basis and averaging loses valuable cognitive information present in individual trials. Tenable conclusions cannot be drawn without all the information present in the EEG – and much of what is removed is removed on account of this so-called 'human bias' having determined what constitutes noise and what doesn't. Perhaps something we see as noise, because of prior human knowledge, may be a key neural marker in understanding something unprecedented about the brain.

Machine learning will be further explained through the dissection of the definition above by Tom Mitchell (1998), in terms of task ("T"), performance ("P") and experience ("E"), in the context of the aims of this research. This explanation translates to the application of deep learning as well. These distinctions will also be explained.

2.5.2 Task ("T")

The task refers to that which the machine learning model needs to do. As examples, this could be a classification task (sorting information into different classes, such as diagnosing a tumour to be malignant or benign); or alternatively a prediction task (using a present state, output the next state as a probability of occurring). Prediction is common in seizure research.

Applying machine learning to the (general) task of EEG analysis is not new. In fact, there has already been great success with real-time seizure detection and the prediction of seizures using and support vector machines (SVMs) and feed-forward networks. Machine learning can be loosely divided into two categories – shallow learning and deep learning. SVMs are an example of shallow learning and the more 'traditional' approach to machine learning. Feedforward networks are an example of deep learning. These real-time results aid doctors in making timeous clinical decisions.

The task of seizure detection is a supervised classification task of separating the ictal and non-ictal phases in the EEG. The most commonly used machine learning algorithm for seizure detection is the support vector machine [SVM] (Fergus *et al.*, 2015, 2016; Johannesen *et al.*, 2016) which is a large margin classifier . The model attempts, through a supervised process, to create as large a margin between the positive (ictal phase) and negative (non-ictal phase) classes. This large margin aims to account for as large variations as possible within the data and between the positive and negative classes. Thereby determining which of the data are most "relevant" or alike to the respective classes (Johannesen *et al.*, 2016). Support vector machines (SVMs) are at their most effective when given hand-crafted, or engineered, features, which is essentially the raw input that has been processed in some way. More information about these features and the SVM can be found in appendix B.1.

The task ("T") of this research is classification; i.e. to classify an action ("left" or "right") using only subconscious EEG data in order to determine whether or not there are features in the subconscious EEG data that can be related to decision-making. In this way, we can assign (or not assign) agency to the subconscious for our decision-making processes.

The machine learning model, or simply model, learns to complete the task through the experience ("E") to which it is assigned.

2.5.3 Experience ("E")

The model will complete this task through a process of *learning* from the data (Goodfellow *et al.*, 2016). The manner with which the ML model learns will be through the experience "E" to which it is assigned. There are different types of experiences to which we can assign our model, such as unsupervised and supervised learning, or alternatively reinforcement learning. Reinforcement learning (RL) involves the deployment of an 'agent' which observes its environment, learns a strategy (termed a "policy") through a self-driven process of trial and error to learn actions, governed by a "reward" and "penalty" system; such as is commonly used in the training of robotics. That said, reinforcement learning is beyond the scope of this research and will not be discussed further.

Supervised learning will used in this research. Supervised learning is the "experience" of the models being given labels to guide its learning. The labels serve as a marker of "correct" or "incorrect" against which the model can continuously compare its initial output estimations. In the beginning, these estimations are akin to guessing, but as the model learns, theoretically its

estimations should edge closer to the 'truth'. The labels are used as a 'truth' comparison, thereby providing supervision to the learning process. This requires the assumption that the data can be separated into pre-determined classes before being given to the model. As in the case of this research, the labels given to the model to guide its learning as to which EEG data belongs to are "left" decision cases and "right" decision cases.

In contrast, with unsupervised learning, there are no labels - i.e. no answer guide / teacher labels to help the model find a pattern. The model learns to distinguish for itself between different classes of data – with minimal guidance regarding the designation of classes and the inherent distinctions between classes. Essentially, we would want the machine learning algorithm to find a way of completing the task, on its own. In this way, the MLA will assign the data to its own designation of classes.

Figure 8 below illustrates the key distinction between the datasets for supervised learning algorithms and unsupervised learning algorithms. In the first graph, with labelled data, there are two pre-defined classes presented to the learning algorithm; whereas in the second graph the dataset is unlabelled, and it will be the task of the algorithm to find a pattern in the data in order to perform something such as classification. This is done by creating a complex function that can separate the classes. In the first case, the model will learn how to identify features in the two classes that will help it make this separation (or, classification) on new unseen data. In the unlabelled, unsupervised case, the model will learn distinguishing features between the data. Dependent on these distinguishing features, it will then separate the data into a representative number of classes, again through the self-guided development of the function.

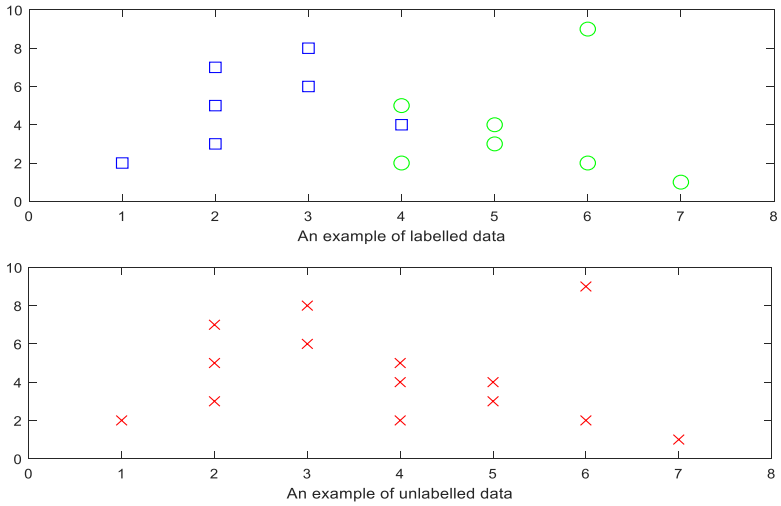


Figure 8: A comparison of labelled datasets (supervised learning) and unlabelled datasets (unsupervised learning)

Only the process of supervised learning will be further explained as it is relevant to this research.

2.5.3.1 How supervised learning models learn

In the experience of supervised learning, the model is given an answer guide which guides its learning. This answer guide is in the form labelled data. In this research, there are two labels: "left" and "right". These labels will be used to distinguish the cases in which the participant chose "left" and in which cases the participant chose "right".

A model consists of a function that tries to learns to find a "best fit" of the data. Best fit can refer to best means of separating "patterns" and/or features representative of - and unique to – each class. The model learns by analysing the subconscious EEG data (in this case) and then output a best estimation of the class to which the data belongs. This estimation refers to the class the model will assign the data (i.e. it intelligently guesses "left" or "right") and is outputted as a probability. This probability can be understood as the model's estimation of how sure it is a piece of data belongs to the class to which it has assigned it to. This estimation is then compared against the actual answer. The actual answer is denoted y. This comparison of the estimation of the answer to the actual answer, is the key feature of supervised learning. This comparison is calculated mathematically using the cost function. The cost function is the mathematical measure of how close the model's output (prediction of the label) is to the actual answer.

Each time the model gives an output, the cost function measures how close it is to being correct. The cost function essentially calculates the distance between the model's 'estimation' and the actual answer. A decrease in this distance indicates the model is able to produce 'estimations' closer to the actual truth. The goal of this part of the process is to have the cost function as close to zero as possible. Further information regarding the cost function can be found in appendix B.2. A cost function of zero would mean the estimation the ML model made is exactly equal to the actual answer - indicating it has done a good job of fitting the data.

As mentioned, at this point the model has developed a basic function to try and perform this task of classifying "left" and "right" EEG data. The model is guided by the cost function as to whether it needs to change its current state of the function or not. This 'changing' of the function is undertaken by parameters. Parameters (not to be confused with hyperparameters which are selected by the programmer) are what the model learns from the data in order to obtain the best fit of the data; to ensure an appropriate estimated output (the model's calculation

of what y could be). The ML model applies a function (the form of which we choose) that takes input training examples x and parameters θ as input. The cost function then assigns a measure of how well the parameters are suited to the task.

Learning occurs through an iterative process of minimising the cost function (reducing the difference between the actual answer, y, and the estimated output (or guess). The hypothesis ($h_{(\theta)}$) is a function parameterised by θ, such that $h_{(\theta)}$ best fits the data. This best fit means that the (estimated) output is as close to the real value of *y* as possible. This will mean there is a minimum error in the model's estimation. The goal is to converge to the (global) minimum of the cost function. This will mean the 'perfect' form of the function has been found, with the optimal parameters. This will allow the ML model to find a decision boundary, or a means of separating these classes (viz. into "left" or "right"). When faced with a new (unseen) example, it will use what it has *learned* from looking at the labelled data and then classify the new data into one of two classes.

However, without a tonne of luck, it is unlikely the ML model will arrive at the optimal values of its parameters (θ), and giving it the minimum cost function following one training epoch and a single updating of the parameters. This process of changing parameters needs to be guided. Assuming the parameters are randomly initiated, the model will need to systematically iterate this process over and over until it reaches (converges to) the global minimum of the cost function. The answer to this is gradient descent.

Gradient descent helps with the goal of minimising the cost function $\big(J(\theta)\big)$ – i.e. guiding the model in adjusting its parameters so as to enable convergence to (ideally) the global minimum. Figure 9 will be used as a reference for the explanation.

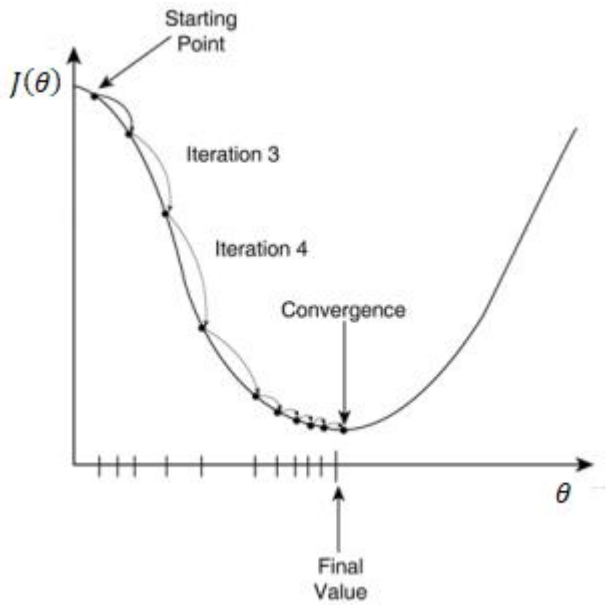


Figure 9: The iterative process of adjusting the parameter (θ) to help the cost function to be as close to zero as possible. Adapted from Mahajan (2018)

The next question is, when does the model update its parameters? How do we test ML model's learnt parameters on data it hasn't seen? It is for this reason the data are split into three groups, viz. training, validation and testing sets.

2.5.3.2 Training, validation and test sets

How do we test the MLA's learnt parameters on data it hasn't seen? Before training, the dataset is divided into a training set, a validation set and a test set (ratio of 70-15-15, for example). The test set is set aside, until such a time we are ready for a final performance check of the model before deploying the model new, unseen data.

The purpose of the test set is to acquire an estimation of the model's performance on unseen data. Testing a model's generalisability is important, i.e. test its performance when faced with unseen data. This is an objective measure of how well the model performs when faced with unseen data.

The model will train on the training set, compare its 'estimation' to the actual answer y. It will then calculate the cost function and then assess the current version of the function on the validation set (Benjamin *et al.* , 2018) – this is a kind of intermediary test to see how the model performs on 'new' and unseen data. Should the cost function be too high, another training epoch will need to be commenced after updating the parameters. This process will guide the updating of the values of the parameters (θ) other.

Only once the model has determined its parameters to be optimal and that the function developed be an appropriate of the data, the model be assessed on the "test" data. This is a test of the generalisability of the model on real-world data and therefore the test set is only used once.

The above explains the process of learning, however, there still needs to be a metric to measure how well the model is completing the task. This is referred to as the performance of the model. Measuring (and improving) performance "P" is key in developing a neural network that can be deployed in the real world.

2.5.4 Performance ("P")

The next logical question as this would be: How do we know the model is performing well? In other words, how do we evaluate the degree to which the model is outputting accurate estimates using its optimised $h_{(\theta)}$? This is where performance "P" comes in: we can guide the model by assessing how well it is performing task "T". In a supervised setting, this is a (relatively) simple task, as we have an output, or "correct answer" to directly measure the performance – either the model is matching the correct answer, or it's not (with a range in between). We can determine the accuracy by which the model is able to classify "left" and "right" movements.

Accuracy is presented as a percentage of how many times the model correctly performed the task overall – that is, the percentage of correct classifications of "left" and "right" movements. The test set accuracy is the value of interest – as we want to see if the model was truly able to learn – and this is tested on unseen data. This value should be as close to 100% as possible. An accuracy in and around 50 % in a binary classification task such as this shows a poor performance. This is no greater than chance – the model simply guessed each round and happened to guess correctly half the time.

2.5.5 "T", "E", "P"

In summary, with Tom Mitchell's definition in mind, machine learning is the process of providing a select model with data and parameters with or without an answer guide (depending on task "T"). Guided by experience "E": the model goes through a process of learning, viz. taking the data, analysing it and finding patterns [and features] that may not have been obvious to human analytical methods (Géron, 2017). This learning process is measured against an appropriate metric that we can use to determine how well the model is performing in its task

(whether it be classification or prediction, formally referred to as regression) [i.e. performance "P"] and from there we can either improve the model or see how it does when presented with new unseen data (viz. how well it is able to generalise to new examples).

The basic principles of machine learning apply in the specific context of deep learning.

2.5.6 The specifics of deep learning

Deep learning is a branch of machine learning. The difference is the use of artificial "neural networks" (not to be confused with neural networks in the brain) which create depth in the model. If we consider a model to be "shallow", this involves a single step between the input and output. "Deep" neural networks (DNN) are essentially a series of interconnected steps between input and output. This creates a "deep" network – with stacked algorithms between input and output.

Essentially, the "deep" in deep learning refers to the depth these networks introduce with the layers through which the input is passed. Figure 10 is an example of a simple deep neural network.

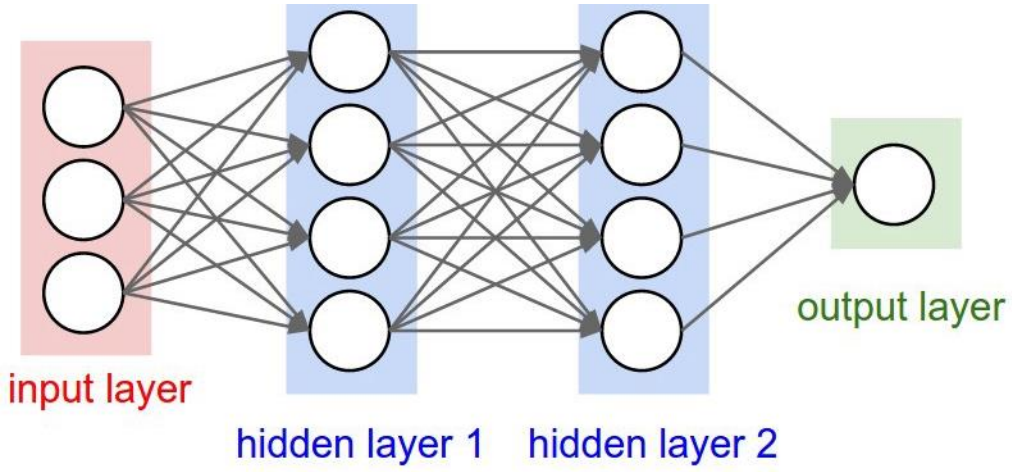


Figure 10: A simple artificial neural network comprising of three input units, two hidden layers; each with four activation units and a single output unit (Karpathy, n.d.)

The output layer will provide the neural network's best 'estimation' (or, 'guess') for the model. The hidden layers are where the learning, described above, takes place. The network comes in with a reticular type formation of interconnecting artificial "neurons" creating an artificial neural network (ANN) with a specific architecture. The architecture of the neural network describes its specific structure: i.e. the number of layers, the number of activation units in each layer and the arrangement of the connections. It is important to note, in literature the term

"neuron(s)" can be used interchangeably with "unit(s)". Hereinafter they will solely be referred to as "unit(s)" in the context of ANN, to avoid confusion with the biological "neurons" of the brain.

Generally, the ANN starts with an input layer, each input unit is then connected to each unit of the first hidden layer, which is then in turn connected to each unit in the next layer (which can be another hidden layer or the output layer, depending on the depth). The hidden layer units are "activation units", which house activation functions (such as sigmoid [logistic] -, softmax -, tanh [hyperbolic tangent] - or ReLU [rectified linear unit] – activation functions); through which the inputs are passed.

The activation functions are non-linear, giving the deep neural network (or, ANN) the ability to learn non-linear functions [this along with the inclusion of at least one hidden layer (van Biljon, 2018)]. Non-linear functions allow the model to learn more complex "patterns" in the data. It is especially important to select activation functions based on careful consideration of their respective properties and potential pitfalls to ensure adequate performance of the chosen model in completing its task. Activation functions are described in detail in appendix B.3.

Each input is passed into an activation unit where the output to the activation unit is mathematically computed and then passed onto the next layer; otherwise it is finally output as the model's 'estimation'. This process is called forward propagation. Following this, the estimated output is evaluated against the actual answer y to assess the cost function. Should the cost function not be optimal (i.e. be as close to zero as possible), the parameters in the deep learning model (neural network) are updated. This is done using a method called back-propagation.

Forward and back-propagation make up the process of "learning" and updating the function so that it finds the best fit of the data in order to complete the task.

2.5.6.1 Selecting a deep learning model

Probably the most logical question to ask at this point is, which model (used interchangeably with the term deep neural network and ANN in the context of deep learning) will be the most suited to tackle the task? There is no shortage of neural networks to choose from. In selecting the model, it is important to consider the task of this research which is to find features in subconscious EEG signals, (i.e. before the moment of conscious awareness), that can predict

what the person is going to do (essentially before they know what they're going to do), allowing us to make inferences about the volition of that action.

It is important to note, that although previously stated that ML brings a fresh face of analysis to the research around the paradigm free of 'human bias's' when analysing the signals and separating noise from useful signals, there is a degree of bias when choosing our models. Models are chosen based on their theoretical components that make them better suited to certain tasks – and therein lies a degree of human bias in that we choose models and their architecture and hyperparameters (for clarification purposes: hyperparameters are selected by the programmer, while parameters are learned by the model) based on how they are understood to be better suited to certain tasks.

Once a model is selected, one needs to also fine-tune the model. That is, to take a basic foundational model build and decide on its architecture, fine-tune the hyper-parameters, and adjust the input data (increase the training set size as already suggested, or tweak the features in a supervised learning algorithm [SLA]). A common challenge in ML (and DL) is achieving sufficiently low generalisation errors (i.e. the model's performance on the unseen test data). Sometimes, the model fits the training set too well, interprets noise as signal and fails to generalise to new examples. The risk for interpreting noise as signal is something to be kept at the forefront of consideration, due to the inherently noisy nature of EEG. This is technically known as overfitting. Overfitting is essentially an overly complicated fit to training data that doesn't generalise to data that is representative, but slightly different (Goodfellow *et al.*, 2016; Ng, n.d.). In the case of overfitting, the function learns the structure of the training data, but this function cannot be applied to new, unseen data. Underfitting is the other end of the spectrum, in that the model is too simple and fails to "capture the data structure" (Ng, n.d.) of the training set, failing to fit the input data. This means it performs poorly on the training set as well as unseen data. This is due to the fact the function doesn't fit the data, it can't apply what it learned about the patterns in the training set to new, unseen data. There are strategies to combat overfitting, such as dropout in the context of convolutional neural networks (CNNs).

The selected deep learning model employed in this research is the convolutional neural network.

2.5.7 Convolutional neural networks

Convolutional neural networks (CNNs) are an example of a deep neural network that has drawn inspiration directly from neuroscience in developing the model architecture (Géron, 2017; Goodfellow *et al.*, 2016). The model is loosely based on the neurons and their receptive fields in the visual cortex. Insights into the inner workings of the visual cortex are the result of the seminal work of Hubel (1959) and Hubel & Wiesel (1959, 1968) on cats and then later on monkeys. Their findings brought forward the theory that groups of neurons each respond to- and process only to a respective portion of the visual field. Consider the visual field (i.e. the scope of the information [picture] passing from the eye to the visual cortex): a group of neurons will each have their own respective receptive field that forms part of the total visual field. In order to obtain the complete picture, these groups of neurons have overlapping receptive fields. Although overlapping, this doesn't mean duplicate information, as each group is responsible for processing different "line orientations" (horizontal, vertical, diagonal etc.) within the visual field. These line orientations are simple features which are added together to make up the visual field and thus we can consider these groups of neurons to be "lower-level neurons". (Géron, 2017)

Some neuron groups will process larger receptive fields that are of a more "complex" nature. These larger receptive fields process more complex patterns by combining the simple features of the lower level neurons. These neurons with large receptive fields are considered to be higher-level neurons. In other words, these receptive fields are based on the outputs of neighbouring lower-level neurons. The simple features are combined into complex patterns, recreating the image in the brain as it was when entering the visual field. (Géron, 2017)

Drawing inspiration from this, the architecture of the CNN was developed. The model architecture itself is further inspired by the Neocognitron introduced in 1980 by Kunihiko Fukushima for handwritten digit recognition.

The principles of how a deep neural network learns, as described above, is still relevant. However, the inner workings of the CNN model are different. The proceeding sections discuss these nuances.

The CNN itself is distinctly unique from other artificial neural networks. In the networks discussed so far, each unit in each layer interacts with each unit in the layer before and after it; with a relevant parameter dictating this interaction (Géron, 2017). In contrast, a CNN has at least one layer of partially connected units. Figure 11 illustrates the basic difference between fully- and partially- connected layers.

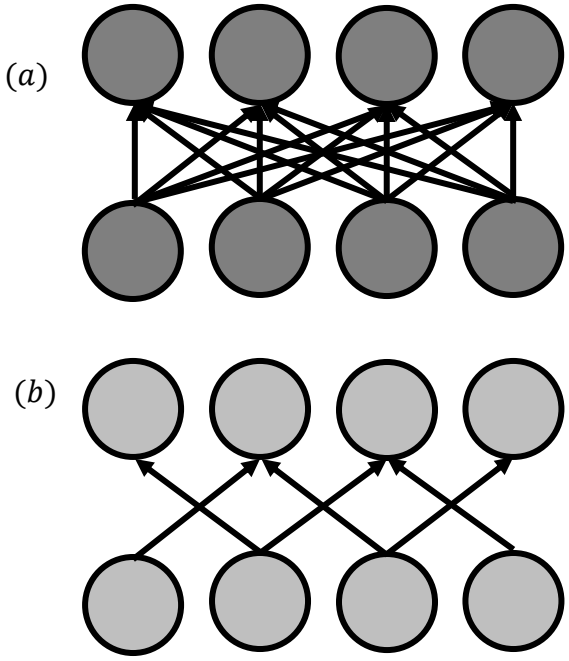


Figure 11: A comparison of fully connected layers (a) and partially connected layers (b). Adapted from Goodfellow et al. (2016)

To gain an intuitive understanding of the work of partially connected units, consider the idea of lower level neurons only processing a portion of the visual field. Each partially connected unit is responsible for a section of pixels in the image (as an example of input). The first layer of partially connected units is responsible for collecting lower level ("simpler") features. The subsequent partially connected layer takes these and assembles them into more complex features. An excellent graphical description of this is available for reference in the article "Visualising and Understanding Convolutional Networks" (Zeiler & Fergus, 2013). The partially connected layers are just one characteristic of the CNN architecture. More characteristics will be discussed next.

2.5.7.1 How do convolutional neural networks come about?

A convolutional layer is an example of a partially connected layer. Convolutional layers are so named as they house the convolution operation making them distinctly unique from other neural networks which employ different mathematical operations. A convolution is a

commutative mathematical operation that acts over two real-valued inputs and outputs a third ("CS1114 Section 6: Convolution", 2013; Goodfellow *et al.*, 2016). Figure 12 introduces the idea of "convolving", or sliding, across an input and producing a smaller output.

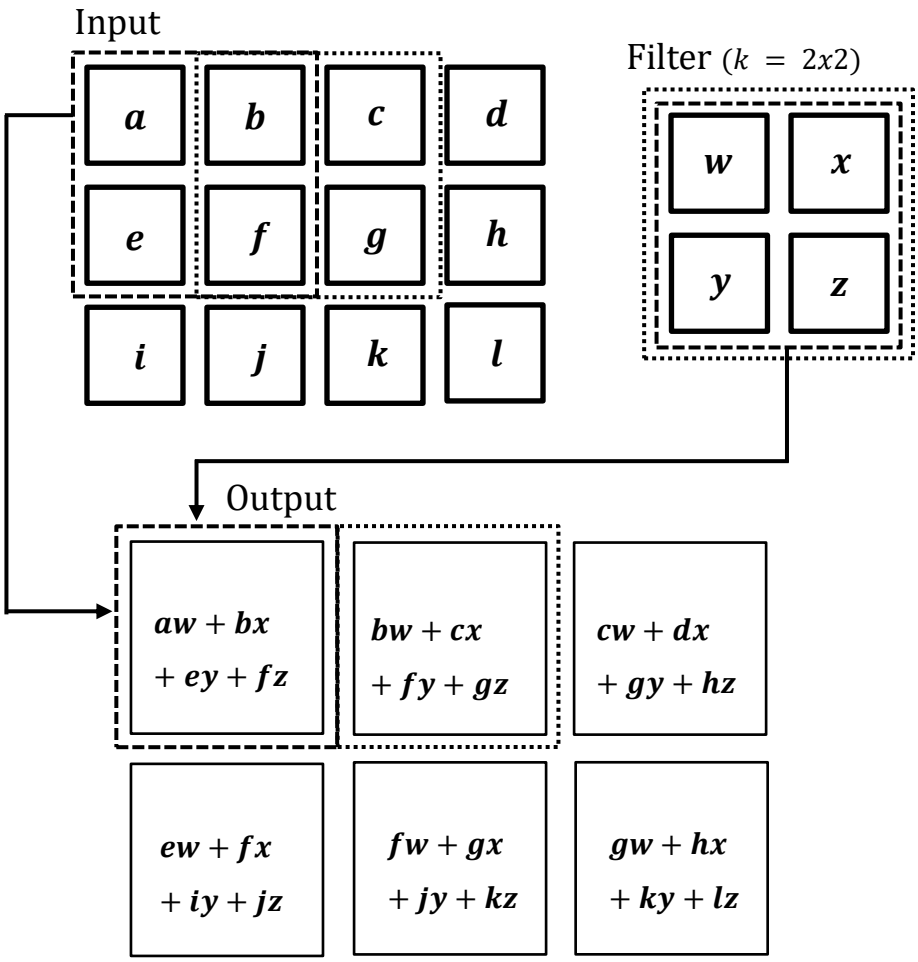


Figure 12: An introduction to the process of "convolving", or sliding across an input. Adapted from Goodfellow et al. (2016)

Inputs to a convolutional layer are 3-dimensional tensors which do not have to be of uniform size to other examples in the training set (Goodfellow *et al.*, 2016). Figure 12 depicts a 2-D input and output, but in reality, these are both 3-D. The dimensions are height, width and depth. It is important to not confuse the depth of the *input* with the depth of the model architecture. The units in each layer are also only connect to a small portion of the preceding layer. (Deep Learning Indaba, 2018b)

As mentioned, the property of the convolution is to slide over the input. The parameters and operations involved in this are defined as follows:

- Filter:
    - Filters are responsible for detecting features in a portion of the input. Filters slide across the entire input. They generally occupy the entire depth of the input that is involved in operation at a specific point in time (Deep Learning Indaba, 2018b)
    - This is a learnable parameter of the CNN
- Kernel size (k):
    - The kernel is synonymous with the word filter. This term is specifically used to define the width and height of the filter. The depth is generally the same as that of the input (Deep Learning Indaba, 2018b; Géron, 2017; Goodfellow *et al.*, 2016)
    - The kernel size can be several orders of magnitude smaller than the input and still obtain good performance. The kernel size used in a specific model is constant throughout that specific model (Goodfellow *et al.*, 2016)
- Stride:
    - The stride defines the distance the filter "slides" across the input with each frame
    - The typical value is one, however two and three are sometimes used (Deep Learning Indaba, 2018b; Géron, 2017)
- Padding:
    - With the reduced number of connections, the output is smaller than the input. Consider Figure 12, where the input has the dimensions 3x4 and the output is of 2 x 3
    - To circumvent this, zero-valued units (known as padding) are added to the height and width to maintain a constant size of the data structure (Goodfellow *et al.*, 2016)

In summary, the operation starts with a "forward pass" – which is to slide the filter (size k) across the full depth, height and width of the input (covered by the filter). This is where the convolution computation comes in: computing the element-wise dot products between the filter and relevant portion of the input. This produces an "activation map" for every spatial position

of the filter. These activation maps are stacked along the depth dimension to produce the output volume. (Deep Learning Indaba, 2018b)

Activation functions (such as ReLU) are also present in depth of the CNN model. These are described in the appendix B.3.

This outlines the operation of a convolutional layer. However, there are another two important characteristics of convolutional neural networks which will be discussed next, viz. pooling functions and dropout.

2.5.7.2 Pooling layers

Pooling layers adapt the output even further (Deep Learning Indaba, 2018b; Goodfellow *et al.*, 2016) in order to aid the learning of complex representations.

Pooling has the added goal of shrinking the input deeper into the network to reduce the computational load, memory usage as well as the number of parameters needed. The pooling layers help make these neural networks more feasible and less expensive (in terms of how much compute is needed) to run.

Figure 13 illustrates the reducing of the size of the input to reduce computational load.

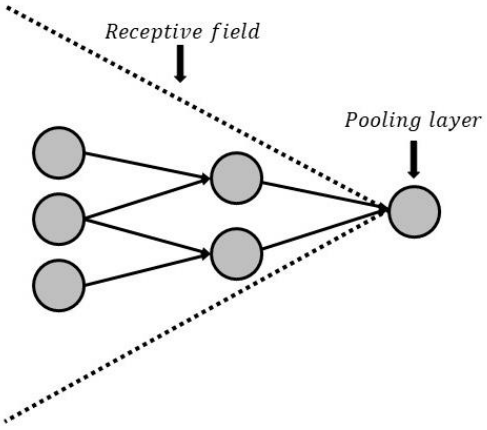


Figure 13: A depiction of how pooling layers reduce the size of the input to reduce computational load

This combining of features involves excluding irrelevant features – only selecting those most relevant to recreating the 'big picture'. Goodfellow *et al.* (2016) defines the pooling functions as: "[replacing] the output of the net at a certain point with a summary statistic of the nearby outputs". This can be seen in Figure 13, where the pooling layer is a single unit that has

combined the features of the units in the layer before it. There is more than one way to achieve this pooling function, but only max pooling will be defined as it is applied in this research

Max pooling is defined by Zhou & Chellappa (1988) as producing "the maximum output within a rectangular neighbourhood". A small window steps across the input (to the pooling layer) and takes the maximum value (most important feature) from each step Although this results in a smaller output, the depth of the output equals the depth of the input. There are no weights (parameters) associated with this function. (Deep Learning Indaba, 2018b).

The next tool to be discussed is dropout. This is a useful tool to avoid overfitting (Pham *et al.*, 2014; Srivastava *et al.*, 2014).

2.5.7.3 Dropout

Dropout is another strategy used to prevent overfitting. Overfitting is an inherent risk when considering EEG data as it tends to be very noisy in nature, and the noise can mask relevant features.

Computers are extremely literal entities. For example, if we consider two images of the same thing with a slight difference for identification, unless each pixel is in the exact same place, the computer at first glance *might* determine they aren't the same. Consider Figure 14 below of two "x" letters: to the human eye, they are essentially the same as we can see past the "noise" or added pixels and we can identify both as "x". However, to the computer, these are two completely unrelated images. This literal logic becomes an issue when dealing with data that is mostly similar with inherent variations at the same time (consider handwriting recognition, or in the case of this research, different EEG signals of two people despite observing the same phenomenon with have different signals due to uncontrollable confounding factors such as drowsiness or hormones). Figure 14 below depicts an example of noise added to an input.

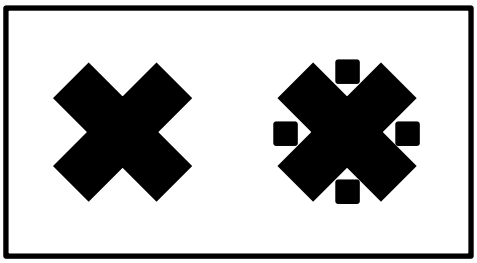


Figure 14: An example of how a network may not recognise two images as being the same due to small changes such as noise.

To avoid this, the model needs a greater degree of robustness in cases such as this. Dropout is a means of achieving this along with a means to prevent overfitting (Srivastava *et al.*, 2014).

At every training step, every unit (except those in the output layer) in the network has a probability p of being completely ignored. This is a hyperparameter generally set to 50% (Géron, 2017; Srivastava *et al.*, 2014). Essentially, what this means is that at each training step the entire network is different. Dropout only occurs during training and decreases the (over-) reliance of any unit on the outputs of neighbouring units. As it is almost virtually impossible for the same network architecture to be sampled twice, each unit needs to be as useful on its own as possible, and the network cannot rely on a few input units only. (Géron, 2017)
Each unit in the model is forced to learn parameters without relying on its neighbours for guidance. In this way, the model is able to learn more relevant features in the data, and not overfit to the noise.
In other words, this prevents the model from learning a kind of hard and fast rule of "seeing this feature, doing that; and then seeing that feature, doing this" [Adapted from the teachings of Allingham (2018)].

Figure 15 below illustrates this idea of different network architectures at each training step on account of dropout, where black depicts a unit that has been 'dropped'. Each of these are of the same model, however they differ between steps on account of the implementation of random dropout. Note that their connections fall out as well, as indicated by the jagged line (Srivastava *et al.*, 2014).

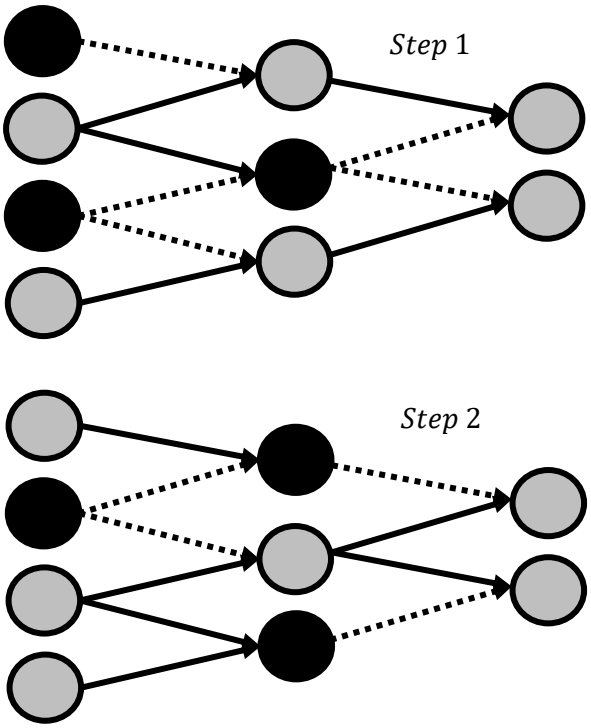


Figure 15: A comparison of two versions of the same model across two frames, with dropout applied

2.6 Summary of the literature review

This chapter introduced the literature around free will, and the neuroscientific pursuit thereof. The theory for EEG and eye tracking was also introduced. EEG and eye tracking formed the basis of data collection in this research. Finally, the theory of machine learning and the branch thereof called deep learning was introduced. The theory behind ML and DL was explained in the context of this research. Deep learning will be the main focus in analysing the subconscious EEG data in order to find predictive neural signals. The chosen deep learning model used in this research was the convolutional neural network (CNN).

The following chapter describes the research methodology, procedures and data analysis pursued in this research.

3 Methodology

Chapter 3 describes the research method and design, the data collection procedure as well as the four components of data analysis conducted: the EEG data pre-processing and ERP analysis, the Libet information, deep learning (for the subconscious EEG data) and the pre-processing and analysis of the eye tracking data.

3.1 Research questions

Can we use machine learning to determine the nature of free will?

Can machine learning be used to classify an action using only subconscious EEG data?

3.2 Study design

The study design is a proof of concept design as we aim to establish new means of predicting movement using EEG signals as well as eye movements discerned from eye tracking data.

3.3 Study setting

The experiment was conducted at the Neuromechanics Central Analytics Facility (CAF) unit in Stellenbosch.

3.4 Study population

Volunteers were recruited from the student body at Stellenbosch University. Recruitment was achieved through the use of advertisements and emails to the student body. Each participant was screened according to the inclusion and exclusion criteria for their eligibility to participate in the study. The initial screening was done using the screening tool can be found in appendix C.1. Following the initial screening, direct questions were asked to confirm eligibility and availability.

No

3.4.1 Sampling

Convenience sampling was the method employed in this research. Similar studies were assessed in terms of their number of participants, in order to set a measurement for the baseline.

The range of the number of participants was 5 to 40 participants. The mean number of participants in these 13 studies was 14.5 participants. (Alexander *et al.*, 2016; Banks & Isham, 2008; Bechara *et al.*, 1997; Fried *et al.*, 2011; Jo *et al.*, 2015, 2014; Libet *et al.*, 1983; Miller *et al.*, 2011; Rigoni *et al.*, 2013; Schultze-Kraft *et al.*, 2016; Soon *et al.*, 2008, 2013; Trevena & Miller, 2010; Verbaarschot *et al.*, 2016)

Thus, in terms of comparison of results, it was determined that a minimum of 30 participants would need to be recruited.

3.4.2 Accessible population

The study population includes individuals over the age of 18 who meet the inclusion and exclusion criteria as set out below.

3.4.3 Inclusion and Exclusion criteria

Table 4 outlines the main the inclusion and exclusion criteria. An unabridged version of the inclusion-exclusion criteria (coupled with explanations and references) can be found in appendix C.2.

Table 4: An overview of the main inclusion and exclusion criteria

Inclusion criteria
Normal, or corrected to normal vision with contacts
Abstinence from alcohol and stimulants 24 hours prior to testing. Abstinence caffeine 2 hours prior to testing (including but not limited to coffee, tea, caffeinated soft drinks)
No prior knowledge of the ‘free will’ component of the study to avoid opinion bias. The study prior to the experiment was described as an investigation into decision-making processes. The specific ‘free will’ component was explained upon completion of the experiment
Exclusion criteria
Impaired vision requiring spectacles. Spectacles cause added reflections which have the potential to confuse the eye tracking system)
The use gel, hair products and/or wet hair on the day of testing
The following is a list of medications used as grounds for exclusion: • Clozapine • Haloperidol

- Tricyclic antidepressant
- Benzodiazepines
- Antihistamines
- Opiate-based analgesics (e.g. Demerol)
- Warfarin
- Antihypertensive medication
- Active chemotherapy or radiation
- Any recreational drug use (including but not limited to marijuana)

The following is a list of medical conditions used as grounds for exclusion:

- Epilepsy
- Cerebral palsy
- Concussion(s) within the last 12 months
- Active CNS inflammation
- Previous traumatic brain injury(-ies)
- Tumour(s) [both present and as well as previously removed]

3.4.4 Data collection information

Data were collected from 21 participants (4 female) from 1 April to 20 May 2019. This is excluding the pilot data collected in March 2019. All participants were right-handed. The age range was 18 to 28 years, with a mean age of 21.1 years. One participant's EEG data were rejected on account of excessive EEG artefacts present within the data.

3.5 Procedures

3.5.1 Measurements and instrumentation

The following equipment was used:

- The electroencephalography was recorded using the BRAIN Products ActiCAP EEG system with 128 gel electrodes
- The eye tracking was recorded with Tobii 4C Eye tracking device with an upgrade key to access the raw data
- The modified Libet clock (Vinding *et al*, 2013) was presented using the PsychoPy2 (Peirce *et al.*, 2019; Peirce, 2007) coder functionality onto the computer screen

3.5.2 Experimental procedure

Once informed consent was obtained, all participants were instructed to follow a procedure similar to the Libet paradigm, with the main diversions from the original paradigm as follows:

- Eye tracking data was collected simultaneously with the EEG data.
- Participants were asked to click a button with either their left or right forefingers at will. This was changed from the original Libet paradigm, on account of one of the objectives including the prediction of the "what" of the freely willed decision. In other words, predict which hand the person will move before they are even aware of their decision. This addition of a choice is important in the context of free will as it allows us to meet two out of three criteria as laid out in Chapter 1 (viz. voluntary, and allowing an alternative path -so that the person was able to act "otherwise" should they choose to do so (Dias & Lavazza, 2016).

3.5.2.1 Requisite lighting conditions

In order to comply with the best practice requisites for EEG and ET data collection ambient lighting was ensured during the duration of the experiment. Dimly lit rooms are best practice for EEG testing. Direct sunlight and artificial light can introduce noise into the eye tracking data (iMotions, 2018).

3.5.2.2 Some nomenclature

In order to make sense of the experiment, it is important to clarify some concepts used to describe the different parts of the experiment described in section 3.5.2.3. Table 5 is included for reference.

Table 5: Nomenclature used to describe the experiment

Concept	Explanation
Trial	One free willed decision – start of the Libet clock to action
Round	A collection of 11 consecutive trials
Experiment	The collection of 100 + trials completed by each participant

3.5.2.3 The procedure followed for each participant

1. The BRAIN Products ActiCAP EEG system with 128 gel electrodes was placed on the participant's head, having measured for the appropriate size.

2. The electrolyte gel was applied so that impedance between each electrode and the participant's scalp was below 10 Ω. The impedance levels for all 128 channels, plus the ground electrode were recorded for later reference.
3. Participants were placed in a comfortable chair, with a computer keyboard placed in front of them. A computer screen was placed 80 cm away from the participant, which is closer than in the Libet paradigm, but on account of the keyboard, the set up mimicked that of the experiment conducted by Banks & Isham (2008).
4. The following eye tracker demonstration was performed for each participant:
    a. The eye tracker "gaze" feature was activated as feedback system for explanation of how the eye tracker works. As the participant moved their eyes, there was a corresponding movement of a "bubble" on the screen.
    b. The participant was positioned so that the eye tracker was able to pick up both of the participant's eyes.
    c. The participant was instructed to checked their position relative to the eye tracker at the start of each round, to ensure the eye tracker was able to record the position of the eyes.
5. The following EEG demonstration was performed for each participant:
    a. The Libet clock was introduced to ensure the participant knew where to focus their attention during a trial– that is at the cross at the centre of the clock (screen).
    b. It was explained that while the Libet clock is in rotation, the participant needed to have both eyes fixed on the screen for the data to be recorded.
    c. A specific demonstration of EEG artefact generation was performed for each participant:
        i. The participant was instructed to perform actions like blinking, giggling and clenching their jaw to be aware of the effects these actions have on the data. These data were not recorded as it was merely for demonstrative purposes.
    d. In the proceeding steps, simple instructions, with not too many restrictions/ expectations were given. This was to avoid introducing potential cognitive artefacts into the data that result from suppressing certain actions like blinking (Cohen, 2014).

i. The participants were informed to not pre-plan their decision and to press “left” or “right” as soon as this decision entered conscious awareness.
ii. The participants were informed there was no time limit and were further encouraged to rest as needed between trials and to reserve longer rest periods for between rounds.
iii. Once each trial was completed, the screen showed a stationary Libet clock. The participants were asked to indicate at what point in time (using the arrow keys to move the dot on the screen) they became consciously aware of their decision to move. In other words, use the interactive Libet clock on screen to record “W”.

This instruction was given to them at the beginning of the experiment, but as specified in the original 1983 paradigm, this was only to be considered at the end of each trial to avoid pre-planning.

e. After each trial, the participant was asked to indicate whether or not there were any mistakes in the preceding trial. Examples of errors included pre-planning and/or forgetting their moment of conscious awareness.

6. Each participant was given the chance to practise and ask questions until such a time they were comfortable with the experimental procedure.
7. The initiation of each trial was self-paced and began only once the participant was completely relaxed.
8. Each participant completed a minimum of 110 trials. Some rounds were repeated in the event of excessive error and/or noise noted during the experiment.

The next section outlines all the data collected during the experiment described above.

3.5.3 Summary of data collected

The following table, Table 6, outlines all the data collected during the experiment. It is presented in the order that the analysis thereof will be discussed in the proceeding sections of this chapter.

Table 6: Data types collected during the experiment

Data type	File type
"W" time	.csv Recorded as an angle relative to the angle of the action
EEG	.vhdr and .EEG files: contains the EEG timeseries information as well as the triggers for event designation .vmrk: contains position information as well as the reference data
Eye tracking	.csv file Records eye gaze information relative to time

3.5.4 Ethical considerations

Ethical approval has been obtained from the Health Research Ethics Committee at Stellenbosch University. Informed consent was obtained from each participant prior to the trial. All information obtained and data collected has been treated as strictly confidential. All data obtained during the experiment is stored under code names on a standard hard-drive. Anonymity has been maintained as there was no follow up (i.e. no need to reference participant names with the code designated. All reference to the participants' personal details, including the code names, are being kept by the principal researcher.

This study has been conducted with the assurance that the basic human rights of the participants were upheld as stipulated by the South African Bill of Rights and Batho Pele principles. The study has adhered to the ethical guidelines and principles of the international Declaration of Helsinki, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research. The data collection involved minimal to no risks for the participants, however, participants were repeatedly advised they were allowed to withdraw at any point without any detriment or consequence. Participants were not compensated financially for their time, in order to ensure true volition in their decision to participate, however a selection of snacks were provided in adherence to their indicated dietary requirements and allergens in order to maintain adequate blood sugar levels during the time intensive and cognitively demanding experiment.

3.6 Data analysis

The following section outlines the data analysis. This was separated into four parts in order to prepare the data to meet the various objectives. An overview of the data analysis is as follows:

- EEG data pre-processing:
    - The EEG data was pre-processed for two reasons:
        - An event-related potential (ERP) analysis was done in order to recreate the RP as found in the original Libet paradigm
        - The data was prepared as input to the convolutional neural network (CNN) used in the deep learning analysis
- Libet information
    - This was investigated in order to determine the average recording for the subjective moment of conscious awareness ("W")
- Deep learning analysis:
    - This involves the input of only subconscious EEG data into the CNN model in order to classify the action "left" or "right"
    - This was done to satisfy the objective of determining the role of the subconscious in our decision-making processes
- Eye tracking data pre-processing and analysis:
    - Eye tracking data was prepared in order to determine the plausibility of using the eyes as a more objective measure to time-lock the moment of conscious awareness ("W")

Each of these aspects of data analysis are detailed below.

3.6.1 EEG data analysis

Before the ERP analysis and being given as input to the machine learning model, the EEG data needs to be pre-processed and prepared. The following section is reserved for the explanation of the manual EEG pre-processing pipeline.

Pre-processing is an important step in EEG analysis. By nature, EEG data is noisy as it picks up more than just the brain's electrical activity (Cohen, 2014). EEG records the electrical activity of nearby tissues (i.e. volume conduction) as well as from non-biological sources such as line noise. The aim of pre-processing data is to remove as much noise as possible. This is to

allow the CNN model to find “patterns” in clean subconscious EEG data that is made up of mostly cognitive components.

Some noise is unavoidable and in the words of Mike Cohen (2014): “One scientist’s noise may be another scientist’s signal”. The idea is to let the CNN model determine what it can from the data, discerning noise from signal itself, with the data having undergone the minimum requisite change. A balance must be found, however, between under and over pre-processing. Raw EEG data will have too high a signal to noise ratio, and the noise may be overpowering. This could mask any important signals or cause the CNN to interpret noise as signal leading to poor performance thereof. EEG data that is excessively pre-processed may result in the inadvertent filtering out (for example) of important cognitive data. Careful applications of pre-processing measures are necessary.

EEG pre-processing was performed using the source software EEGLAB with its variety of plugins (Delorme & Makeig, 2004). EEGLAB is a toolbox and graphical user interface (GUI) based in the MATLAB programming environment (MATLAB 2019a, The Mathworks, Inc.).

There are standardised automated pipelines such as the Harvard Automated Pre-Processing Pipeline for Electroencephalography [HAPPE] (Gabard-Durnam et al., 2018). However, HAPPE relies on the automatic algorithms of EEGLAB and leaves little room for intuition regarding changes to the data. This is especially important to consider as these algorithms were designed with specific assumptions in mind – i.e. for more ‘pure’ neuroscientific analysis such as ERPs; and not for machine learning. Machine learning requires an inherently different approach. It is for these reasons that the “manual” pipeline as implemented by An et al. (2019) and Anders et al. (2018) was followed to account for these differences in assumptions and approach.

A checklist outlining the entire process (which was completed for each participant's data during pre-processing) can be found in appendix D.1. The entire EEG pre-processing pipeline is summarised in Figure 16, following which the pre-processing of EEG is described in full.

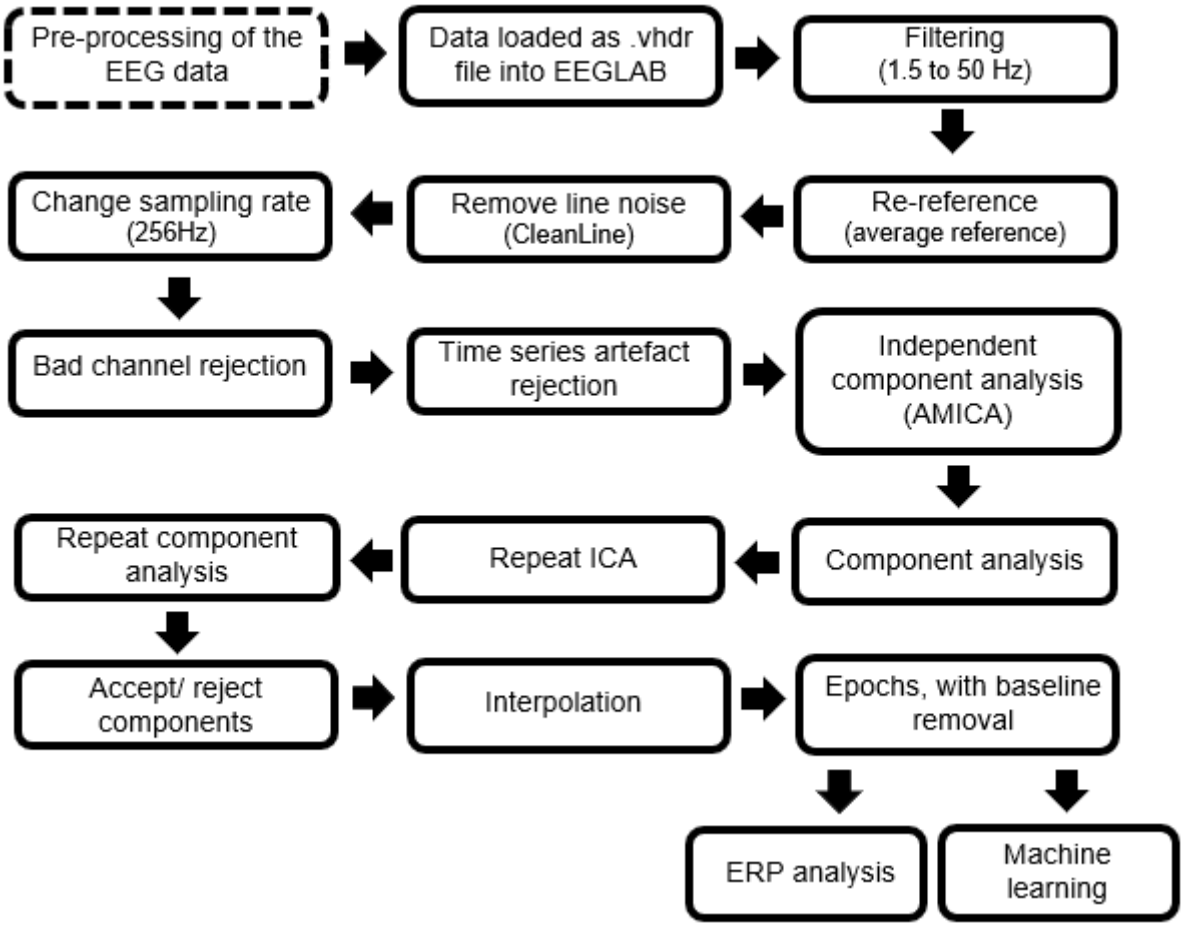


Figure 16: A flow diagram of the pre-processing steps for EEG data

3.6.1.1 Overview of the process

There are four main steps in removing noise from the data. These are listed as follows:

- Filtering:
    - To ensure only the frequencies of interest are included (i.e. from delta to gamma frequency waves)
- Bad channel rejection:
    - Excessively noisy channels can skew all other channels measured relative to it as the chosen referencing technique is average referencing
- Artefact removal:
    - This is process of removing non-stereotypical, non-recurring artefacts

- Independent component analysis (ICA):
    - This is process of removing stereotypical, recurring artefacts

Each of these steps are described in detail below.

3.6.1.2 Filtering

Filtering is necessary to increase the signal to noise ratio by removing as much noise as possible. Typically, noise emanates from external sources such as line noise, or excessive motor (muscle) activity and these are filtered out by excluding certain frequency and/or frequency ranges from the data. EEG signal and noise can be interlinked (e.g. eye blinks are summed on top of the neural processing signal) and require more intricate methods such as independent component analysis (ICA) for removal, as opposed to simply excluding certain frequency ranges from analysis. ICA will be discussed in this chapter.

A broad spectrum of EEG frequency waves was included so as to avoid assumptions as to the range in which we may find the neural markers. Thus, in order to meet the aim of preparing the data for the machine learning: the high and low pass filters were 1.5 and 50 Hz respectively, to include gamma rays (which also form part of the cognitive recording spectrum).

To clarify:

A high pass filter can only be applied to continuous data, to block signals so low in frequency they carry more noise than signal (Schomer & Lopes da Silva, 2011). The ideal range is between 1-2Hz, for adequate ICA decomposition and to avoid distorting the signal (Winkler et al., 2015). The high pass filter was applied at 1.5 Hz, so as to include the majority of the delta waves (δ) [these typically occur between frequencies of 0.1 – 3.9Hz].

A low pass filter blocks signals so high they carry more noise than signal (Schomer & Lopes da Silva, 2011). In this pipeline, the low pass filter was applied at 50 Hz to include a part of the gamma (γ) frequency range [40 -100 Hz]. Applying the low pass filter to a frequency greater than 50 Hz was found to introduce too many artefacts in to the data. This distorted the cognitive data too much for analysis and increased the model's affinity to overfit to non-cognitive noise.

Despite these filtering bounds of the high – and low pass filters, it was still necessary to apply the CleanLine filter due to the potential for residual line noise frequencies in the data.

Line noise is typically from sources such as AC power line fluctuations, medical equipment (hospital setting) or fluorescent lights and is found at 50 or 60 Hz depending on the country

where the data are collected [although these exact values cannot be guaranteed (Mullen, 2012)]. It is assumed that the South African line noise is at 50 Hz, as per convention. CleanLine, an EEGLAB plugin will be used to remove line noise (Mullen, 2012).

3.6.1.3 Sampling rate

The online sampling rate was 500 Hz. In accordance with this pipeline the data was down-sampled to 256 Hz (An *et al.*, 2019; Anders *et al.*, 2018). According to the Nyquist theorem, the sampling rate needs to be at least twice as high as the highest frequency of interest. The highest range of gamma waves (without filtering) is 100 Hz, so assuming we would like to investigate the highest range of brain waves, a sampling rate of 256 Hz satisfies this requirement.

At this point the data were ready for re-referencing and observation for further removal of noise.

3.6.1.4 Referencing technique

The choice of the type and location of a reference is important as a reference chosen in an electrically non-neutral area might distort the topographic data. This is on account of each electrode's electrical potential calculated in relation to the reference electrode's potential. (Teplan, 2002)

Different referencing techniques can be found in appendix D.2. The average reference was used in this pipeline (An *et al.*, 2019; Anders *et al.*, 2018) as there is an even distribution of electrodes across the entire scalp for each participant. However, before the average reference was applied, the data at Cz (the online reference) was included into the matrix.

3.6.1.5 Rejection of bad channels

The data are inspected across the time series for bad channels initially, and then artefacts. 'Bad' channels are removed when they carry more than 70 % non-stereotypical noise. A bad channel can potentially result from the following causes:

- Insufficient contact of the electrode with the scalp (due to poor cap fit or extremely thick hair)
- Faulty electrode(s)

- A gelled electrode where, over the course of the single experiment the gel dries up and the impedance gradually increases to be greater than 10 Ω

In cases such as this, the noise far outweighs the signal (low signal to noise ratio) and filtering the noise out would essentially leave no signal. In such an instance it is best to remove the channel entirely. Another reason to remove the channel is to avoid skewing the entire data matrix with noise as the referencing technique is to average across all channels. However, removing more than 13 electrodes (~10 % of the total number of channels) electrodes would require rejection of the participant data as a whole. This rejection criterion is necessary due to the later pre-processing step "Interpolation" where a representative signal is generated from surrounding electrodes to 'fill in' the rejected electrode (Gabard-Durnam *et al.*, 2018). Removal of 13 electrodes would mean there is not enough 'true' EEG data left to generate valid EEG data in the removed channel spaces. In order to generate truly representative EEG data, there needs to be enough reference data.

This step is performed by viewing the channels in the timeseries, and any channel that has more than 70 % noise (i.e. <30% EEG data) was rejected. If any channels are removed, interpolation will need to be performed following the second ICA analysis so that representative data can be inserted in the spaced of the rejected channel.

3.6.1.6 Rejection of artefacts in the time series

Manual rejection of artefacts involves the analysis of the time-series channel data. Artefacts need to be rejected to limit the number of non-cognitive components being assigned to a component sources in the ICA is to identify the maximum number of independent *cognitive* components as possible.

When analysing the time series, data are rejected across all 128 channels, even if the artefact is only present in one channel. Careful deliberation to distinguish stereotypical from non-stereotypical artefacts should be employed when rejecting data, to mitigate the amount of data loss through rejecting time series across all 128 channels.

Stereotypical artefacts have a typical, recurring waveform. Examples include eye blinks and electrocardiography (ECG) which is picked up when an electrode is placed over a pulse. These will be allocated a component in the ICA. Removal of each stereotypical artefact in the time-series would result in too much data being lost across 128 channels, thus it is necessary to remove these artefacts in the ICA analysis.

In contrast, non-stereotypical artefacts do need to be removed in the time series as these are atypical and non-recurring. If left to the ICA, these artefacts may take up an entire component channel, and only occur once or twice in the entire timeseries of the experiment. ICA component sources are limited by the number of channels in the EEG electrode set up – so this would reduce the potential number of cognitive components that can be isolated. The idea is to maximise the number of independent cognitive components, and minimise the number of atypical non-recurring artefact sources in the ICA source analysis.

In short, stereotypical and recurring artefacts are left for removal in the ICA component analysis. Non-stereotypical, non-recurring artefacts are removed in the time series.

Good practice when removing artefacts in the time series is to maintain the "zero line". As data are selected across all 128 channels, it is important to note that the points that are "cut" are at the zero line – anything removed in a peak or trough will create a non-stereotypical artefact when the data are concatenated again.

During the removal of artefacts, it is possible to delete the entire trial if the person coughed or spoke (as examples) during the trial. It is necessary to keep record of the time stamps used for rejecting in the time series. This is made easy with EEGLAB's "eegh" function which keeps a log of all transforms made to the data.

A baseline was set for participant rejection. When removing artefacts, it is possible to also delete entire trials. Participants that had more than 25 % of their Libet clock cycles rejected, were excluded entirely. Following this, the final stage in removing artefacts was implemented, viz. ICA.

3.6.1.7 Independent component analysis (ICA)

ICA decomposition is employed in order to separate components within the continuous EEG signal. It has been mentioned, but it is important to remember: EEG is not exclusive to neural processes. EEG records a multitude of electrical potentials which sum on top of one another. These signals can be mixed within a signal channel, or within a subset of channels within a specific region. In order to remove stereotypical artefacts such as eye blinks, muscle twitches and residual line noise, these components need to be separated out from the neural processes, as well as from other neural processes. Other neural processes refer to continuous processes that aren't related to the task at hand.

ICA aims to find *statistically (stochastically) independent components.* This means the occurrence of one signal does not affect the probability of another signal occurring. For example, in the EEG recording regions close to the eyes (a) channel(s) may pick up both neural processes and eye blinks. The presence of eye blinks doesn't exclude the probability of there being neural processes as well in the same signal (eye blinks sum on top of neural processes), making these statistically independent. To build further intuition around this: two time series are maximally independent when their signals (or, "waves") are maximally distinct (different) from each other (Makeig & Onton, 2011). Further information regarding the theory of independent component analysis can be found in appendix D.3.

Another key factor in isolating independent components, is the dipolarity of cognitive sources of data. True EEG sources, within the brain are dipolar; whereas EMG or electrical noise are unipolar (only positive or negative). The positive-negative nature of independent sources allows the algorithm to distinguish between different source components. (Delorme *et al.*, 2012)

Once the components are identified, they are assessed according to the following criteria:

- Source localisation
- Strength
- Frequency

A detailed description of each of these criteria can be found in appendix D.4.

All non-cognitive components were rejected. This process was repeated twice as some components may only contain one small section of artefactual EEG data, and the idea is to optimise the number of functional components. In this event, the short time-series section with the artefact were removed. The ICA is run the second time on the pruned dataset, without the single artefact. The second ICA leads to the identification of new independent components.

Once a component is rejected, these components are subtracted from the EEG matrix.

It is acknowledged this is an inherently subjective process. However, strict record of the component analysis under each of the feature headings listed above (see appendix D.4.) was kept for each participant. Each component was accepted or rejected based on a thorough analysis of all five features in conjunction.

3.6.1.8 Data segmentation into EEG windows time-locked to the action

The EEG data were segmented into windows (or epochs) for event-related potential (ERP) analysis and as input to the deep learning model (i.e. the CNN model).

The duration of the two windows of consideration was fixed to 1.5 and 2 seconds, respectively before the trigger. Each window was time-locked to a specific trial event. In this case the time-locking events will be the "left" and "right" decisions. The window extended to 500 ms after the action, to ensure the trigger label was preserved. This trigger label was important as it served as the label necessary for supervised learning algorithms.

Both chosen window durations of 1.5 seconds and 2 seconds were to ensure adequate inclusion of subconscious EEG data. Following this, the windows are used as input for two separate data analyses, viz. the ERP analysis and the CNN model.

3.6.1.9 ERP analysis to reproduce the readiness potential

The ERP analysis was specifically performed to meet the secondary objective of reproducing the results of the original Libet experiment. Successful reproduction of was necessary to ensure the data collected is of high quality and that the experiment conducted is in fact an accurate representation of the original paradigm.

The 1.5 second windows were averaged across all trials and all participants using EEGLAB's ERP function. An ERP analysis was performed for the EEG data relating to the "left" and "right" decisions.

The next step in the EEG analysis is the determining the subjective moment of conscious awareness ("W").

3.6.2 Libet information analysis

This section deals with the method of extracting the information relevant to the original Libet experiment, viz, the subjective report of the moment of conscious awareness of a decision to move, "W".

3.6.2.1 Recording the subjective moment of conscious awareness ("W")

The Libet information refers to the time stamp of the subjective report of conscious awareness ("W") of the decision to move, as well as the time taken to complete the action ("M"). These data are recorded as angles in a .csv file format during the experiment. A Python script was

used to extract this information, calculate the times of "W" and "M" respectively which are then written to a new .csv file. All code used in this research can be found within the following GitHub repository: *https://github.com/SMHall94/Masters_EEG_FreeWill.git*.

In order to convert the relative angles for "W" and "M" into seconds, the angle between the respective "W" and "M" need to be determined. To convert the angle to seconds the following equation is implemented using a Python script:

$$Reaction_{angle} = M_{angle} - W_{angle}$$

Equation 3.1

$$Reaction_{milliseconds} = Reaction_{angle} \times \frac{2560}{360}$$

Equation 3.2

This information is used to calculate the average reaction time:

$$Reaction_{seconds} = \frac{Reaction_{milliseconds}}{1000}$$

Equation 3.3

The reaction time is calculated on an individual trial basis for each person. A grand average is then calculated using all of these values.

The reaction time can be used to mark "W" on an individual trial basis:

$$W_{time} = M_{time} - Reaction_{seconds}$$

Equation 3.4

All data are included in calculating the times for "W", unless the participant marked a trial as erroneous. During the data collection, the participant is instructed to indicate any errors in reporting the moment of conscious awareness of their decision. An error can refer to the self-reported "W" not being recorded during the experiment and/or the participant indicating they couldn't remember the moment of conscious awareness.

This step is important to create a baseline for the separation of subconscious and conscious EEG data, so that only subconscious EEG data is fed into the machine learning model. The deep learning method is discussed next.

3.6.3 Deep learning analysis

In order to explain the deep learning analysis, it is important to re-iterate the primary aim of this research. The aim is to determine the role of the subconscious in our decision-making process. This means only subconscious EEG data will be used to classify the action "left" or "right". A high classification accuracy using only subconscious EEG data would suggest there are features in the subconscious data relating to the decision, indicating a large role of the subconscious in our decision making.

The following flow diagram (Figure 17) outlines the deep learning process.

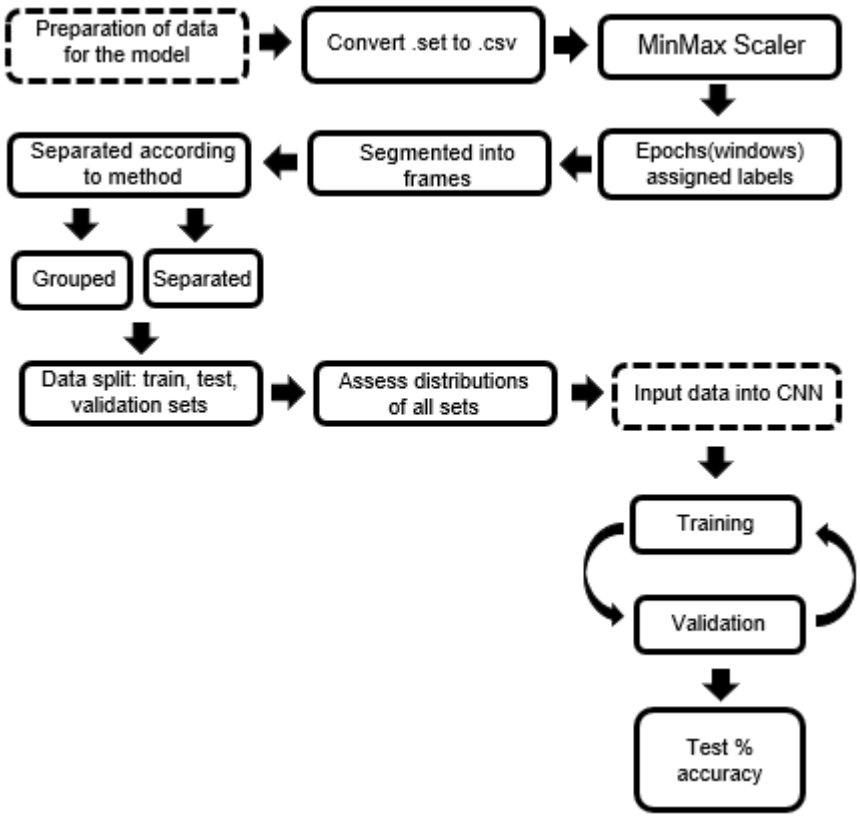


Figure 17 : A flow diagram of the deep learning process

Before explaining the data preparation and model implementation in detail; a few terms need to be re-clarified for easy reference. Table 7 is used for the re-clarification of terms.

Table 7: Nomenclature used to describe the machine learning method in this research

Term	Concept
"W"	Self-reported moment of conscious awareness of a decision. In other words, this marks the moment of separation between the subconscious and conscious
"M"	Moment the action ("left" or "right") occurs
L/R	Reference to the decision "left" or "right"
Window	This is the EEG matrix segmented for analysis Formally known as an epoch in neuroscience, but will be referred to as a window henceforth, to avoid confusion with a training epoch in machine learning
Frame	This is the percentage of the window of EEG data fed as input into the CNN model

3.6.3.1 Selecting the duration of EEG windows

The duration of the windows were 1.5 seconds and 2 seconds. In the original Libet paradigm, the RP arose 350 ms before conscious awareness, creating a baseline for the earliest point before "W" that the prediction using subconscious data should occur.

In order to include the labels relating to the action ("right" and "left") embedded in the EEG data, an additional 0.5 seconds of data is included after the action. It was important to ensure preservation of these labels as they form the basis of supervised learning.

3.6.3.2 Data preservation

This is process of taking the MATLAB .set file, converting it to .csv and then the final data preparation for input into the CNN model. This is explained in detail below.

3.6.3.2.1 Convert .set files from MATLAB to .csv files

This step converted each individual epoch into to a .csv file. (comma separated value file, which can be opened in Microsoft Excel). This conversion involved the time series frequency and amplitude being changed into the respective corresponding numeric values.

3.6.3.2.2 MinMax scaling

The MinMax scaler scales the data to have ranges between 0 and 1; reducing the standard deviation as well as the effect of outliers in the data. MinMax scaling was an important step, as it made the data more suitable to be read by the model.

The equation used in MinMax scaling can be found in appendix E.1.

3.6.3.2.3 Ensuring only subconscious EEG data was fed into the CNN

The next step was to convert the .csv files into Python's "numpy" arrays. These were then entered into a dictionary with the key referring to their class label ("left "or "right)". These labels were very important as this is a supervised learning task.

In order to segment the data for input into the CNN model, the following method is applied as illustrated by Figure 18:

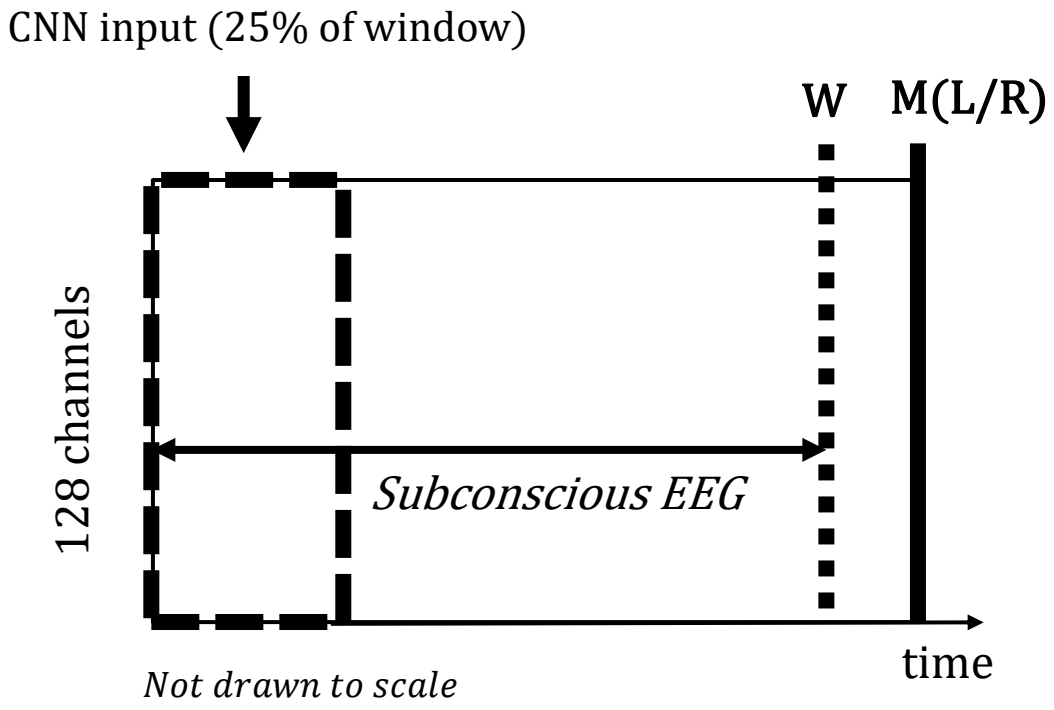


Figure 18: The method of segmenting the data for classification of the action before conscious awareness

This method of creating frames ensures the data fed into the CNN model were only from the subconscious portions of EEG. If order to determine the role of the subconscious in decision, the model needed input only from this subconscious section, to avoid possible overlap of the conscious and subconscious.

The inclusion of the 500 ms after the action was taken into consideration when calculating the 25 % of data used as input (see section 3.6.3.1). Table 8 includes the timestamps of the data fed into the CNN. It is important to notes these time values are before the action

Table 8: The timestamps of the data input to the CNN

Duration of the windows (ms)	Duration of the frame (ms)	Time stamps of the frame (ms)
1500	375	1500:1125
2000	625	2000:1375

Once each window was segmented into the relevant frame, the data were then separated based on the methods two different methods.

3.6.3.2.4 Splitting the data and confirming the distribution

The data are split into three parts: the training set, the validation set and the test set. The split is 70% training data and 15 % for the validation and test sets respectively. The test set is kept aside until the very end, as its purpose is to test the deep learning model's ability to perform on unseen data. The following section describes the two methods used to obtain the split as described above.

Two methods of splitting the data are proposed, viz. grouped according to individual trials, or separated by participants. These methods are described in Table 9:

Table 9: Description of the data separation methods

Method of separation	Explanation
Grouped	All the data were first grouped together and then split randomly into the training, validation and test sets.
Separated	The data were separated by participant. Some participants (and all their respective trials) will be randomly assigned to the training set, and the validation and test sets respectively.

These two methods were employed to further assess the generalisability of the model. This is due to the inherently variable nature of EEG. In the 'grouped' method, there is representative data across all three datasets. The 'grouped' method split all the data across all three sets, ensuring representative data in each subset. Representative data in this context refers to EEG from each individual being representative of their own individual brain presentations. This makes it easier for the model to generalise as it was exposed to data representative of each participant in the training and validation sets before seeing the same representative data in the test set.

In the 'separated' method, there was no representative, or similar data in all three subsets. Some participants' data were placed in the training set, different participants' data were in the validation set and then an entirely different subset of participants were in the test set. essentially, the CNN model trained on some participants, and then tested on entirely new participants. The model still achieved near perfect accuracy. This means the features identified by the machine learning model are not specific to one person, but can generalise to EEG data from a different person, despite the inherent differences as a characteristic of EEG.

The final stage in splitting the data is to check the distributions of each label (i.e. "left" an "right") in each subset. The model assumes a normal (consistent) distribution across all the sets (train, validation and test). This is assessed before each deploying the model through a series of histograms.

Should the distribution be imbalanced (number of trials in one class outweighing the other) additional steps would need to be considered such as down sampling the larger class by randomly deleting trials.

At this point the data were ready for input into the convolutional neural network.

3.6.3.3 The convolutional neural network

The model used in this research was a convolutional neural network (CNN). The architecture was based on that of Schirrmeister, *et al.* (2017). The model of Schirrmeister, *et al.* (2017) was developed specifically for EEG data, by adapting the state of the art AlexNET CNN used for image recognition. The model is coded using the high level neural network platform, Keras (Chollet, 2018). Keras is written in Python (compatible with Python 3.6) and is compatible with both central processing- and graphics processing units (CPUs and GPUs respectively). The basic architecture is that of a four-layer convolutional neural network, with one dense

layer. The ReLU activation unit was chosen to be used in the hidden layers, with softmax used in the output layer. A dropout probability value of 50 % was used (i.e. each unit had 50 % chance of being included in that relevant iteration). Hyperparameter adjustments were only made to the model using the only the validation accuracy. The model is only deployed on the test set once for each window size. The test set is fed into the CNN model only once – at the very end. The test subset is akin to data the model will encounter when released into production (for example) and the test set results shouldn't inform any hyperparameter tuning, as this can result in the model simply overfitting to the test set as well. This test accuracy is recorded for both methods (separated and grouped).

The final analysis to be discussed, before presenting the results, is that of the eye tracking.

3.6.4 Eye tracking analysis

Chapter 2 provided a background to the inner workings of the eye. This section will explore the (pre-) processing of the data and how the theory can be applied in practice.

As explained in the experiment set up: the eye tracker is recording the entire duration of the Libet clock on the screen, up until the person makes their decision. Regardless of the *conscious* focus; the eye tracker records all – including the eyes' movements of what the person is aware about their eyes.

The eye tracking data are collected into a .csv file (comma separated values which can be opened in Microsoft Excel) using a Tobii 4C Gaming Eye Tracker (with upgrade key). The triggers recorded from PsychoPy are synced with the start of the Libet clock as well as to mark their decision (left or right). This, along with the system time are recorded in the.csv. The "gaze" data, or eye movements are recorded as tuples, with a corresponding x and y co-ordinates for each eye. These are recorded at a sampling rate of 90Hz.

3.6.4.1 Proposed hypothesis and method

The inclusion of eye tracking in the Libet paradigm is to determine a more reliable method of marking the moment of conscious awareness of a decision. With reference to the original Libet experiment (Libet *et al.*, 1983), the retrospective and subjective reporting of "W" has come under heavy criticism. Evidence has already shown this report to be vulnerable to manipulation and ultimately inaccurate (Banks & Isham, 2008; Lau *et al.*, 2007).

The advantage of the eye tracker is its non-discriminate recording of data: it records all movements, including those we are not consciously aware of. The eyes are able to act independently of conscious control. This analysis sought to investigate whether the moment of conscious awareness had a corresponding action in the eyes. This corresponding movement can be recorded by eye tracking this change and ultimately find an objective and definitive marker for "W".

The hypothesis is that there will be an eye movement relating to the decision entering conscious awareness. The instruction to the participants was to take note of the position of the clock at the moment they are consciously aware of the decision to move.

The basis of this hypothesis is that in each trial there is a common event just before "M" – a decision enters conscious awareness. This event occurs irrespective of whether or not the decision arose in the subconscious. This decision, regardless of its origin, always results in the action "M". Identifying a recurring eye "event" occurring just before the action could theoretically be inferred as a marker for conscious awareness. This is what this section sets out to investigate.

Pupil dilation has been shown to correspond to changes in attention and awareness. However, pupil dilation has a slow response to changes in cognition and is therefore not an appropriate avenue in the context of this investigation. The reason for this is that the trial time is also very short and does not allow enough time for reliable and time-sensitive pupil dilation measurements. Gaze data, and specifically fixations, are more appropriate as having ascertained eyes respond to changes in cognition (Blignaut, 2009) – this is a good point from which to investigate when and where the brain took in information.

This brings us to the focus of this particular investigation; i.e. the "gaze" data. This is the relative change in pupil position over time – in other words, the velocity. Velocity can be related to saccades have a velocity greater than or equal to $60^{o}.second^{-1}$ (Holmqvist *et al.*, 2011).

Before describing the analysis method in detail, the following describes the pre-processing of the raw eye tracking data.

3.6.4.2 The case of the missing data

In terms of eye tracking data, the most important consideration is eye blinks and participants changing position which result in missing data. Another consideration is mitigating noise during data collection.

Measures are taken to reduce noise during data collection, such as dimming the lights – however, noise is inherent in origins of biological and physical signals. Left unmitigated, the noise might distort signals and therefore change the results computed from them. (Juhola, 1991)

When considering noise, it is important to have an intuition as to what is happening in terms of the data points and which movements relate to a physiological source of noise such as an eye blink, or rather a non-physiological source of noise. An example of a non-physiological source of noise would be the eye tracker losing contact with the eyes, and the algorithms calculating no data, or inaccurate data.

When recording, the eye tracker software reports NaN (not a number) when one or both of the eyes are not able to be picked up by the hardware. This can occur due to blinks, looking away, position shifts or physical obstructions. In the absence of these, data it is difficult to make inferences about the eyes and the correlative brain activity.

Unlike EEG, where the eye blinks sum on top of cognitive data, there is now a blank space without any information – and using a kind of interpolation will only create false data that are not representative of changes in cognition (Olsen, 2012)

According to the Tobii Technologies strategy (Olsen, 2012) – it is only suitable to fill in gaps that are shorter than that of an eye blink, which in the case of this research would be less than four missing data points. As we cannot make inferences about cognition relating to eye movements when there is no data available, this information is not included in analysis. In other words, no interpolation is done in the event of an eye blink.

3.6.4.3 Identifying eye events

An eye event was identified as a saccade of more than $60^{o}.second^{-1}$. Saccades of less than 3 points were rejected (Holmqvist *et al.*, 2011). All these are events were timestamped, and labeled Eye-time.

3.6.4.4 Method of analysis

Eye tracking was included in this experiment as a means of determining whether or not the 'eyes' can be used as a more valid and accurate measure of "W" – i.e. the moment of conscious awareness. The general idea is to find an eye movement / activity that can be related to the moment a decision enters the conscious awareness. In other words, an eye movement that can accurately time-lock the point in time the person decided to move "left" or "right".

In order to do so, we need to make an assumption: we assume the decision is made within 500 ms of executing the action. The person was instructed to press as soon as they became aware of the decision to move. Thus, in order to correlate an eye movement to the conscious awareness of a decision, it needs to at least 500 ms before the action.

This assumption of the decision occurring within 500 ms of the action is also based on research explicitly investigating reaction times. Average responses to auditory stimuli are between 140 – 160 ms and average responses to visual stimuli are 180 – 200 ms (Thomson *et al.*, 1992). However, a reaction to a prompted stimulus introduces a confounding factor to the reaction time. The speed of the stimulus reaching the cortex needs to be accounted for. Visual stimuli can take 20 – 40 ms to reach the cortex, while auditory stimuli can take 8 – 10 ms (Kemp, 1973). Despite there being no stimuli to have responded to in this experiment, these studies confirm 500 ms to be a plausible duration of consideration in terms of a reaction time to a spontaneous, unplanned decision.

In order to complete this method, all eye activities are time stamped. This event will be denoted Eye-time. The action is still denoted "M". The threshold refers to the window in which the eye movement (Eye-time) needs to fall in order to be correlated to a decision. Figure 19 illustrates this method. The double-sided arrow indicates the area of interest for Eye-time. Trials were excluded if the total length of trial was less than 500 ms.

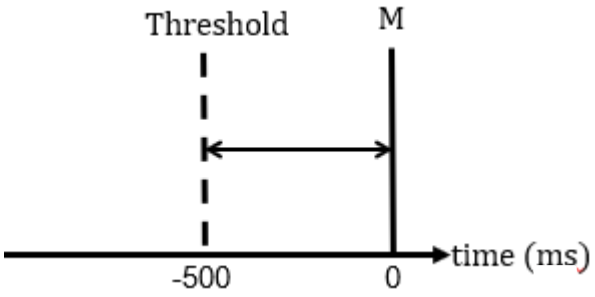


Figure 19: The method used to determine the relationship between Eye-time and "W"

In summary, the method of analysis is based on the assumption the difference between the moment of conscious awareness of a decision to move and the resultant action not greater than 500 ms. The following logic was used to determine the number of eye events that occurred at least 500 ms before the action:

$$If\ eye_{time} > (M - 500\ ms): FALSE$$

$$If\ eye_{time} \leq (M - 500\ ms): TRUE$$

The number of "true" events will be compared to the number of "false" events. In the event of more "true" events than "false" events, this would be indicative of a correlation between the eyes and the moment of conscious awareness of a decision made.

A further comparison was done to determine the influence of the decision "left" or "right" on the occurrence of the eye event (Eye-time). The null hypothesis was that the decision would have no influence on the eye event.

3.7 Summary of methodology and data analysis

Chapter 3 outlined the methodology of this research. A proof of concept design was used as the aim was to investigate new means of investigating the role of the subconscious in decision making using EEG and machine learning, as well as investigate a more objective means of time-locking the moment of conscious awareness ("W"). Data were collected from 21 participants recruited from the student body on Stellenbosch campus via convenience sampling. Data collection took place from 1 April until 20 May 2019. All four methods of data analysis have been explained. The data analysis involved EEG data analysis, determining the information relevant to the Libet paradigm (viz, "W"), the inputting of the EEG data into the CNN model and finally the eye tracking data analysis. The following chapter presents the results of this research.

4 Results

Chapter 4 presents the results of this research. The results are presented in the order of data analysis.

4.1 Recreation of the readiness potential (RP) through ERP analysis

A secondary objective of this research was to reproduce the results of the original Libet paradigm – i.e. the readiness potential. As a refresher: the RP is a rise in neural activity 550 ms before an action has taken place and 350 ms before conscious awareness. Libet and his team took this rise in neural activity before conscious awareness as proof the sub-conscious is acting independently of the conscious and concluded that there is no free will. The readiness potential is the average signal of all the trials at a single channel within the motor and or frontal cortices. For this study, the main point of interest is Cz – the most cephalad point of the EEG cap. The first step, in this research, in terms of data analysis was to reproduce this result. This first figure, Figure 20, is the original result of the Libet experiment – and is the result of averaging out and smoothing across 40 trials and 5 participants.

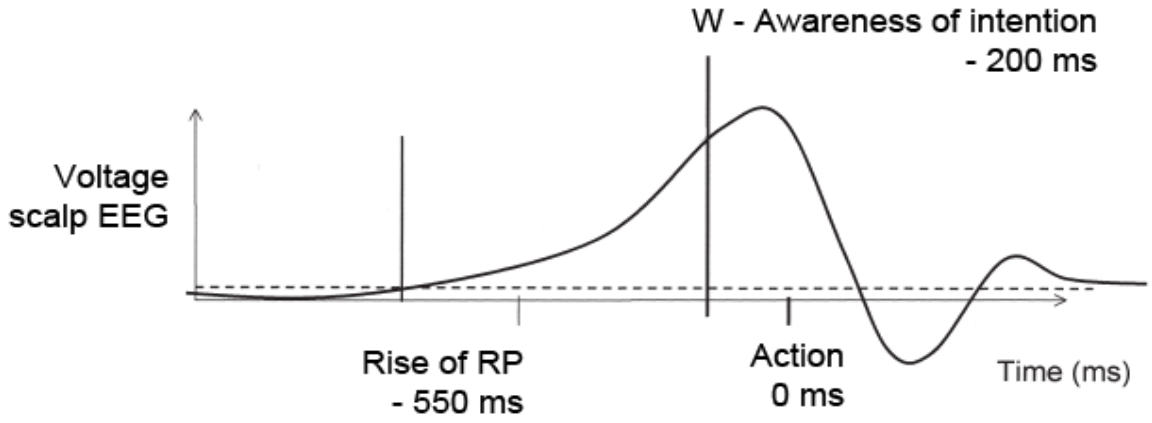


Figure 20: The original result from the experiment in 1983 by Benjamin Libet and his team (Doyle, n.d.)

The following graphs present the results. There are two graphs, one for the "left" decision and the "right" decision. Figures 21 and 22 are the product of averaging out over more than 50 trials in each paradigm across 20 participants. No smoothing has been applied and these are from the Cz channel (the area responsible for executive motor function). The zero line indicated the moment of conscious awareness. In this study, the subjective report for the moment of conscious awareness was found to be 108 ms before the action. In Figures 21 and 22, we see

the rise in neural activity before this mark. This rise in neural activity also occurs 200 ms before the "W" time in the original experiment.

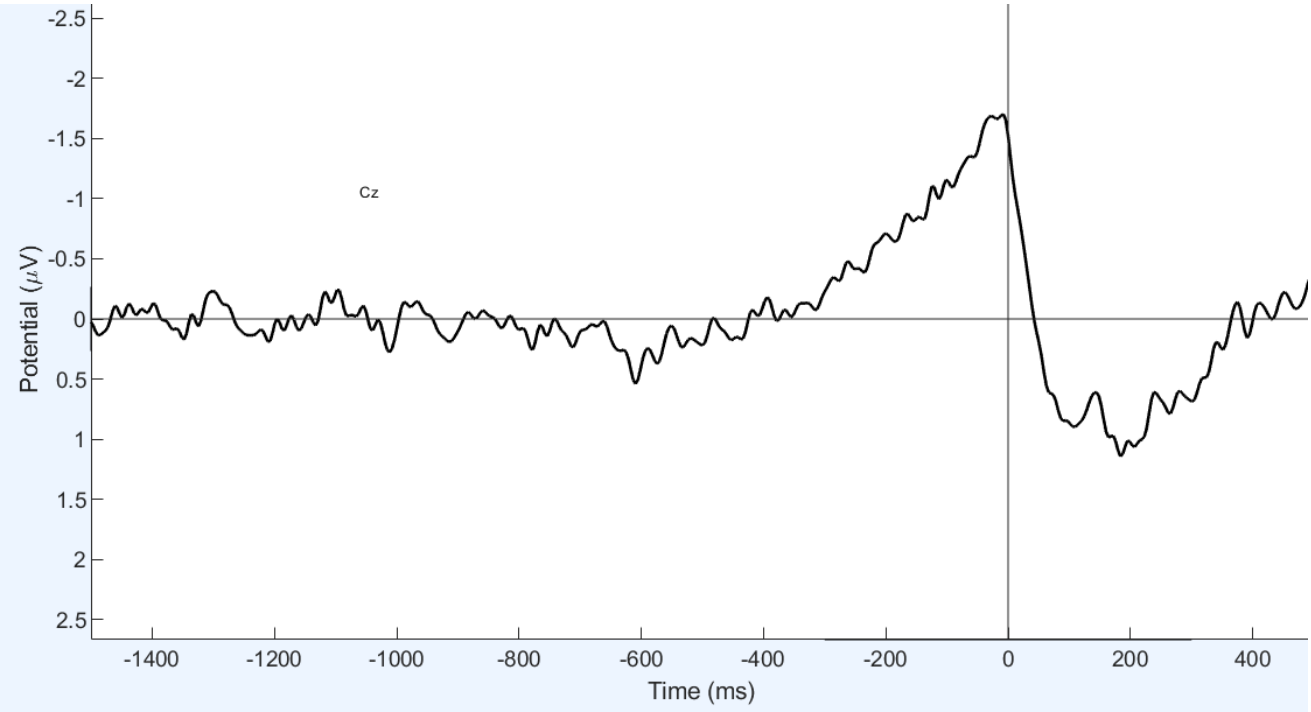


Figure 21: The "left" decision from the Cz channel

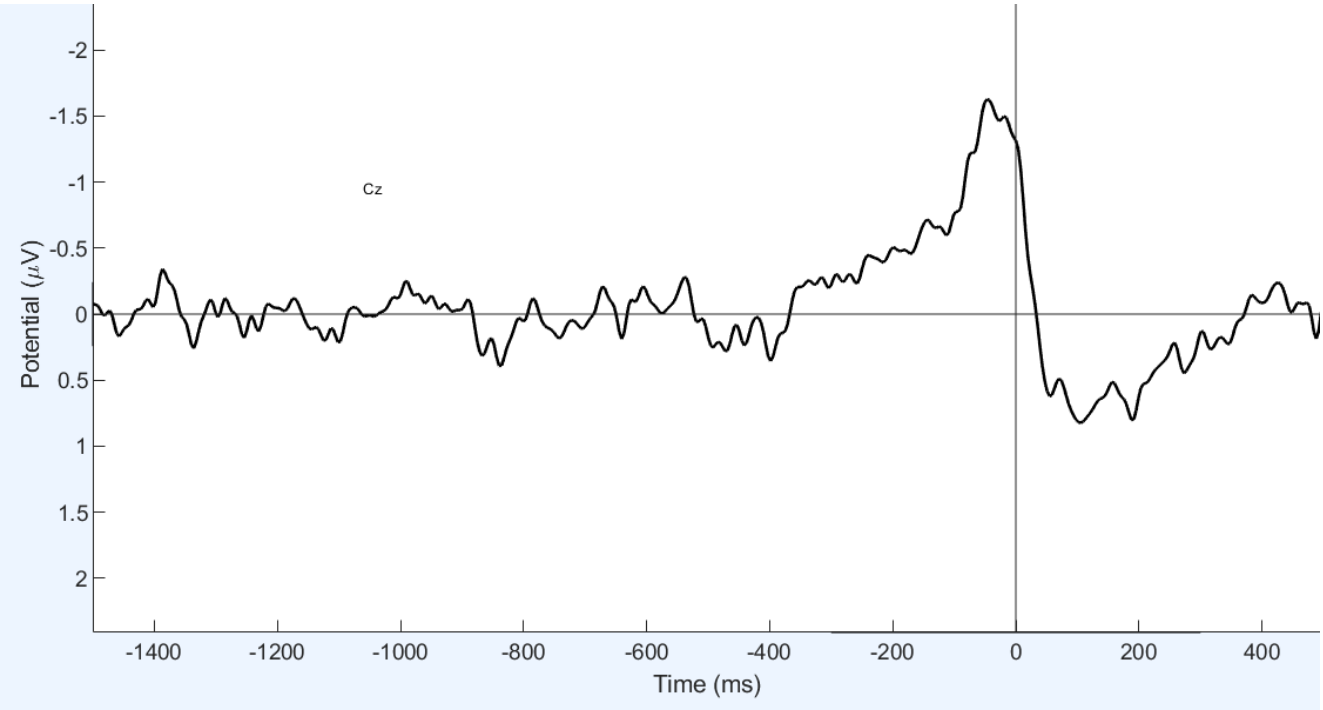


Figure 22: The "right" decision from the Cz channel

This result is not isolated to the Cz channel, but the premise of this objective was to observe the RP at this channel. While successful in reproducing the results from the original experiment, lest we forget that the original work, although groundbreaking, was not without its flaws. A main objective of this research is to address these flaws. The RP is not available on an individual trial basis, and only takes a limited amount of information, and it is difficult to ascertain whether or not this potential is the cause of the action. Machine learning will be used to look at all 128 trials individually, and classify an action before conscious awareness on an individual trial basis.

4.2 Libet information

As per the original experiment, each participant was asked to report the moment they became aware of their decision to move “left” or “right”. These values were, as expected, flawed, as some people reported their moment of conscious awareness to be after the action.

The average reaction time (that is the time between the subjective moment of conscious awareness and the action) was found to be 110 ms. In other words, “W” was found to occur, on average, 110 ms before the action, “M”.

The time obtained for the subjective moment of conscious awareness was used to ensure only subconscious EEG data was fed into the convolutional neural network.

4.3 Deep learning with the convolutional neural network

As described in chapter 3, the data were separated according to two different methods for input to the machine learning model. The model was able to classify with almost perfect accuracy the action “left” or “right” as early as 1.3 s before the action took place, and 1.2 s before subjective awareness of the decision to move. The result for each of these methods are presented in Table 10.

Table 10: Results of the test accuracies

Method	Time before the action "M" (milliseconds)	Time before conscious awareness "W" (milliseconds)	Test accuracy (%) of the CNN model
Grouped	1000	890	99.43
	1375	1.265	99.70
Separated	1000	890	99.69
	1375	1265	98.80

Figures 23 and 24 below illustrate the results more clearly by comparing the time stamps of subconscious activity in this research and in the original Libet experiment's RP. The RP was taken as proof of subconscious control and subsequently criticised for being used as proof of there being no free will. The legend for these figures can be found in Table 11.

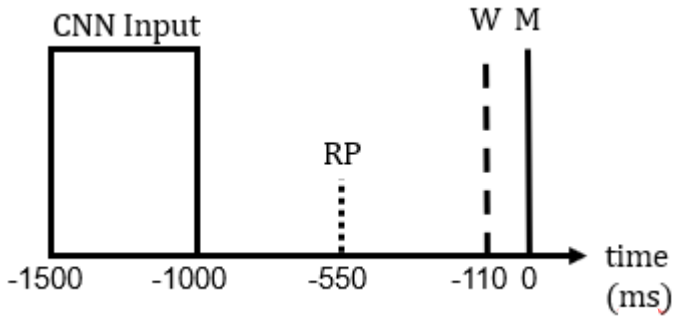


Figure 23: The comparison of the original Libet results and the results of this research with a window of 1500 ms before "M"

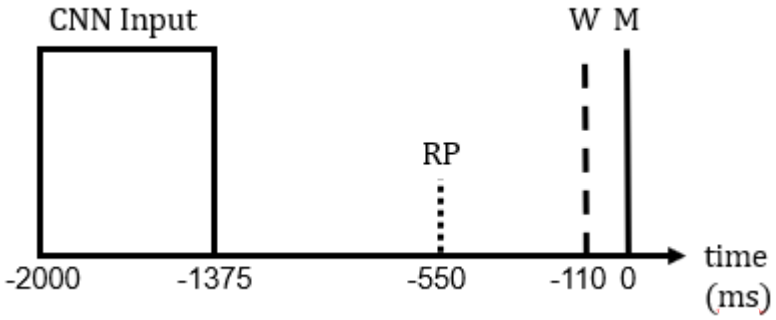


Figure 24: The comparison of the original Libet results and the results of this research with a window of 2000 ms before "M"

Table 11: Legend for Figures 23 and 24

Key	Explanation
CNN input	This is the frame of subconscious EEG data fed into the model
RP	This marks the rise in subconscious neural activity termed the readiness potential in the original Libet experiment
W	This separates the subconscious from the conscious, as per the subjective reports collected in this experiment.
M	The action resulting from the decision ("left" or "right"

The final stage of data analysis involved investigating a more objective measure of the moment of conscious awareness, i.e. "W" using eye tracking.

4.4 Eye tracking

The percentage of "true" events compared to "false" events was 14 %. As a result, no inferences can be made regarding the correlation between eye tracking and the moment of conscious awareness. Figure 25 depicts the results and Table 12 provides the legend for Figure 25.

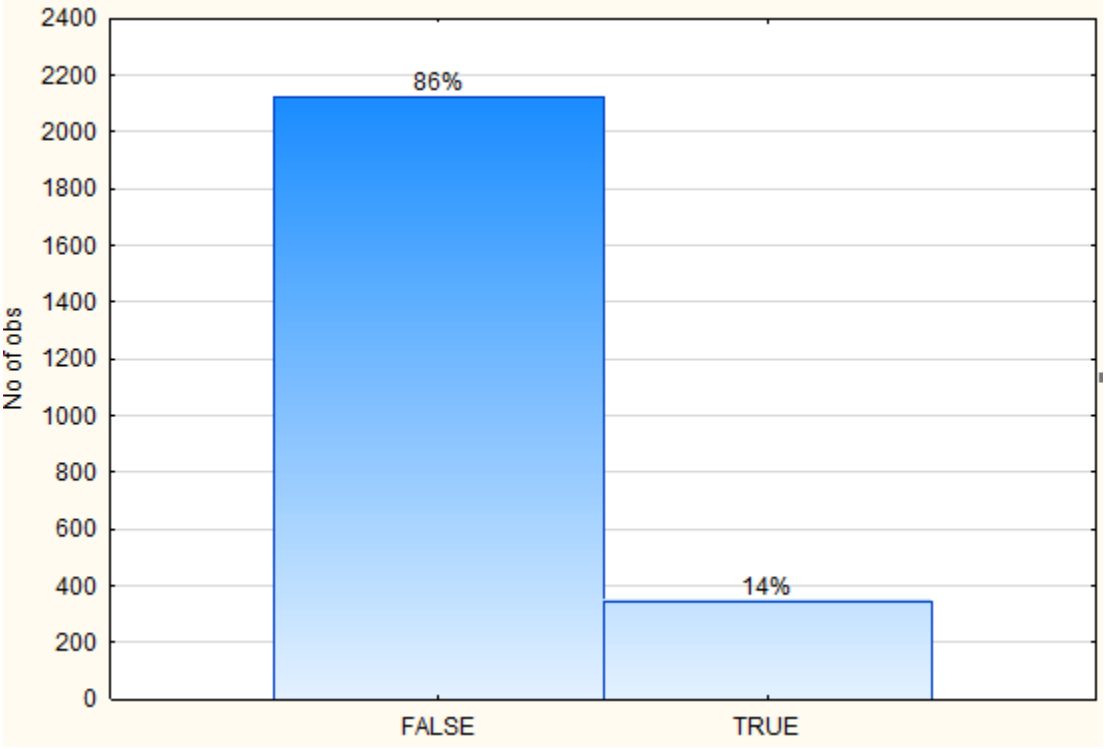


Figure 25: The percentage of eye events within the window of 500 ms before the action "M"

Table 12: Legend for Figure 25

Key	Explanation
No. of obs	Number of trials included in the analysis
FALSE	Eye-time was earlier than 500 ms before the action
TRUE	Eye-time was within 500 ms of the action

The second part of the analysis included determining whether or not the decision "left" or "right" had any influence on Eye-time occurring within the window duration of interest. Figure 26 presents these results. The graph on the left is the "right" decision, and the graph on the right is the "left" decision. Table 13 provides the reference for Figure 26.

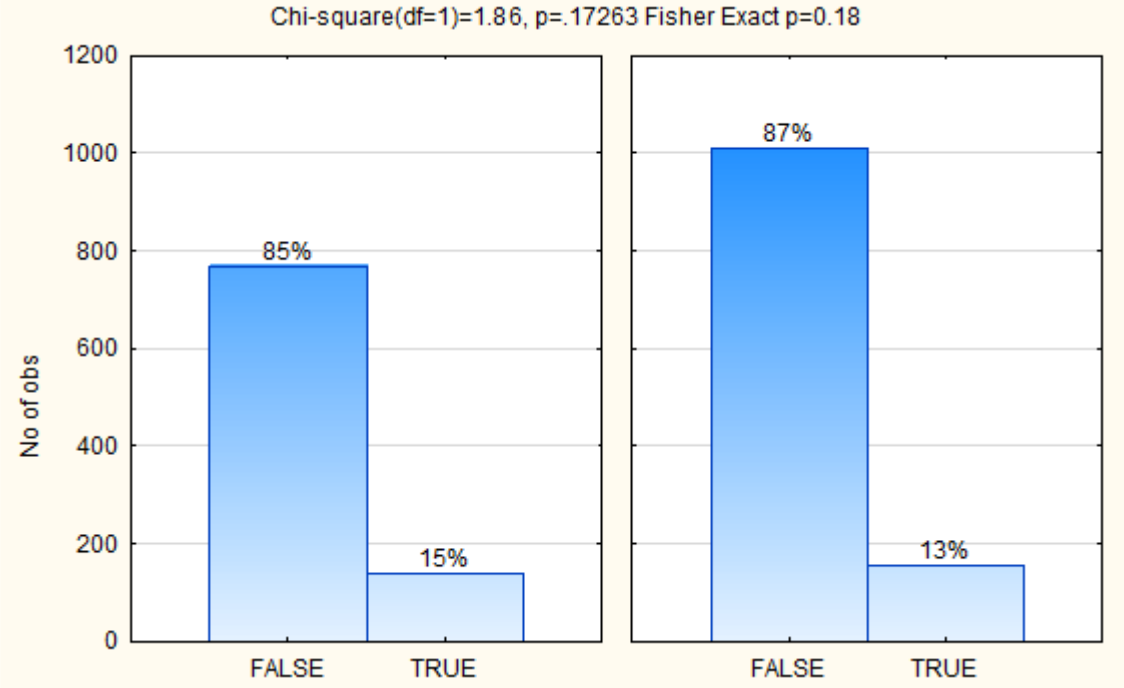


Figure 26: The results of the investigation of the effect of the decision "right" or "left" on the occurrence of the eye event (Eye-time)

Table 13: Legend for Figure 26

Key	Explanation
No. of obs	Number of trials included in the analysis
FALSE	Eye-time was earlier than 500 ms before the action
TRUE	Eye-time was within 500 ms of the action
Graph on the left	"Right" decision
Graph on the right	"Left" decision

The result of this analysis is that the decision of “left” or “right” has no influence on the event of Eye-time. The p-value of 0.17 is not non-significant.

4.5 Summary of results

The results of this research include the successful recreation of the readiness potential for both the left and right decisions. This result proved the validity of the data collected in this research and confirms the original Libet paradigm was successfully reproduced. The subjective report of “W” was confirmed to be imprecise and the average time between conscious awareness and action was 110 ms. The CNN model achieved near perfect accuracy in classifying an action “left” or “right” using only subconscious EEG data. There were no correlations found between the eyes and the moment of conscious awareness. The decision “left” or “right” also had no effect on the occurrence of any events in the eyes.

The following chapter presents the discussion of the results, the limitations of the research and recommendations for future research.

5 Discussion

Chapter 5 presents the findings of this research and the implications thereof. A summary of the limitations of the research is presented, as well as recommendations for future research.

5.1 Introduction to the discussion

Developing a firm understanding of the nature of free will is important to understanding the underlying and fundamental processes of our brains. Understanding free will links to an understanding of moral responsibility and who or what should be assigned agency for our actions. In 1983, Benjamin Libet and his team of researchers introduced the Libet paradigm. This experiment was an attempt to bring empirical evidence to support the debate about free will which previously centered around philosophy and discussion. Although the rise in neural activity before conscious awareness (i.e. the readiness potential [RP]) was taken as proof the subconscious is ultimately in control, the Libet paradigm has faced much criticism.

This main criticism was that the experiment involved a single finger movement – and this cannot be considered anything more than an urge. A free choice requires options, as well as consequence to be considered more than an automatic reaction (Soon *et al.*, 2008). There is the specific requirement of being able to do otherwise in the exact same situation (Dias & Lavazza, 2016). The timing of the RP was also considered insignificant as it occurred only milliseconds before the action as well as being the result of averaging out over many trials. A third, but important criticism, is the recording of the moment of conscious awareness ("W"). This retrospective, subjective report has been proven to be inaccurate and vulnerable to manipulation (Banks & Isham, 2008; Lau *et al.*, 2007).

This research attempts to address these criticisms. A prediction was made using a convolutional neural network, up to 1.3 seconds before the action. This action involved a choice between "left" and "right". This research suggests the notion that there is no free will, under the definition that the conscious self is not in control of our decisions. Agency is assigned to our subconscious and with that in mind, this lends more support to the camp that believes there is no free will.

5.2 Application of deep learning to the Libet paradigm

The first step in analysing the data was confirming that the data collected in this research was representative of the data from the original Libet experiment (Libet et al., 1983). Obtaining this result was necessary to prove the validity of the experiment conducted in this research. Successful reproduction of the readiness potential (RP) at the Cz electrode for both the "left" and "right" decision was achieved. In confirming the validity of the EEG data, further comparison can be made in terms of addressing the criticisms around the Libet paradigm.

In both instances, the RP starts 400 ms before the action which is relatively later when compared to the RP of the original Libet paradigm (Libet et al., 1983). However, this is still 200 ms before the original "W" time, and 290 ms before the "W" of this research. Some differences can be noticed between the RPs (of this research) for the "left" and "right" decision. The overall shape is similar of all three graphs (original and new). The RP for the "right" decision is slightly smaller in amplitude than that of the "left" decision. These differences could be attributed to the hand dominance of the participants (all participants were right-handed), in that the right-handed movements are more "natural" therefore requiring less electrical potential to proceed. This would require the adoption of the "stochastic accumulation" model of Schurger *et al.* (2016) which suggests that all neural activity needs to cross a certain threshold before presenting as an extraneous action. The threshold for "right" movements in a right-handed person might have a lower threshold to cross. However, focused research into this question is necessary in order to make conclusive inferences. These differences in the amplitudes of the RPs also indicate inherent differences in the data characteristic to each decision, which supports the CNN's ability to differentiate between "left" and "right".

The RP produced in this research is subject to the same criticisms raised by Fried et al. (2011), Miller et al. (2011) and Trevena & Miller (2010). Causal relations between the RP and the decisions would need to be investigated in order to conclude whether or not the RP is the cause of the decision and not merely the product of general brain activity and/or an artefact of the clock. However, the investigations around the RP should cease, as it is an average of brain activity and not available on a single trial basis. It can be argued it is a redundant source of information around neural activity. Therefore, it is recommended that future research rather focus on determining causal relations on a single trial basis to better understand the dynamic nature of the brain.

It is for that reason the primary objective of this research was to accurately classify a decision (left or right), by employing a supervised learning deep neural network using only subconscious EEG data as the input. This was achieved using a convolutional neural network with a test classification accuracy of ~ 99 % between 0.8 and 1.2 seconds before conscious awareness (i.e. up to 1.3 seconds before the action). This is near perfect accuracy was achieved on an individual trial basis - an excellent indicator of the model's performance.

Furthermore, the CNN model was able to predict each individual action made by the participant. This result not only evidences that decisions arise in the subconscious, it is able to do this on an individual trial basis using data that is more representative of the brain's function. This improved representation of brain activity is important, as it maintains the complex and dynamic nature of the brain. With each decision, our brains present a little differently, and these differences are preserved in the data inputted to the CNN model. The success of the CNN can be attributed to there being features relating to decision-making more than a second before conscious awareness. This is an improvement on the RP, as it only presented a few hundred *milliseconds* before the action and conscious awareness, which doesn't account for significant subconscious involvement. Although this research did not manage to predict as early as in the experiments of Soon *et al.* (2008, 2013), it is still able to predict at least a second earlier than the Libet paradigm. Soon *et al.* (2008) achieved accuracies of 60 % up to seven seconds before conscious awareness. It is speculated we can achieve these accuracies earlier as well with deep learning. However, some participants in this research decided to move very quickly and as a result there weren't windows of 7-10 seconds from each participant to input into the CNN model. Further, EEG is also generally more accessible and cheaper than fMRI, as used in the experiments of Soon *et al.* (2008, 2013). EEG is safer and less invasive than the method of using depth electrodes, further improving on the results of Fried *et al.* (2011).

Assessing the generalisability of these results was a secondary objective of this research. Despite the model achieving near perfect accuracy, it is difficult to determine the degree to which the model will perform on the general population. This is especially true on account of the EEG being inherently variable and therefore generally difficult to generalise from one person to the next, as well as the small sample size used in this research. It is for this reason that two methods were employed in splitting the data as input for the model, i.e. 'grouped' and 'separated'. The second method, the 'separated' method is relevant in the determination of the generalisability of the model. In the 'separated' method, some participants' data were placed in the training set, different participants' data were in the validation set and then an entirely

different subset of participants were in the test set. Essentially, the CNN model trained on some participants, and then tested on entirely new participants. The model still achieved near perfect accuracy. This means the features identified by the machine learning model are not specific to one person, but can generalise to EEG data from a different person, despite the inherent differences in characteristic of EEG. The success of the 'separated' method indicates the features are generalisable to a degree, however, the research is still limited by the small sample size. Future research should include larger samples, with a greater representation of general, healthy populations as well as representative sample of unhealthy populations with central lesions.

Another major criticism was addressed through the addition of a choice, as the RP has been argued to be no more than an urge (Soon *et al.*, 2008) or the product of unspecific neural activity (Trevena & Miller, 2010). In investigating free will, as it is used in the context of this research there is the explicit need to have been able to do otherwise, given a rerun of the exact same scenario (Dias & Lavazza, 2016). This follows on from the work of Soon *et al.* (2008, 2013) in that the participants were given an explicit choice in each trial. This method is still limited to a degree, in that there is no consequence (a third requirement of a free choice) and this experiment doesn't allow for replication of a real-world scenario such as that introduced in the work of Maoz *et al.* (2017). Further research into the addition of consequence is necessary.

The success of the CNN model gives conclusive evidence that there are features in the subconscious EEG data relating to a decision. These features are identifiable on an individual trial basis, unlike the readiness potential of Libet *et al.* (1983) and are specific to either decision ("left" or "right"). The features are present in the subconscious with the participants having had the ability to do otherwise – which aligns with the definition of free will used in this research (Dias & Lavazza, 2016). This further confirms the agency for our decisions belongs to the subconscious. This supports the conclusions of Fried *et al.* (2011); Libet *et al.* (1983); Soon *et al.* (2008, 2013) and provides evidence to the notion that there is no free will, under the definition that the subconscious has primary control over our decisions. This result has put machine learning's foot solidly in the door of contributing to the empirical neuroscientific field of the free will debate. Despite the results not being conclusive in terms of the debate, this supports more evidence for the school of thought that the conscious self is not in control, and that the subconscious plays a role in our decision-making.

The final criticism that was addressed was the subjective report of "W". The reliability of the report of "W" was analysed in this research in order to address the secondary objective of

determining the precision with which the person is able to subjectively report their moment of conscious awareness ("W"). The average time between the action, "M" and "W" found in this research was shorter than that of the original Libet paradigm [110 ms compared to 200 ms] (Libet *et al.*, 1983). While human error is the most likely cause in light of the inconsistent reports of "W", a timing of "W" as 110 ms before the action is plausible as a measure for the time taken to make a decision and then act on it. Explicitly studied reaction times (difference between thought and action) by Thomson *et al.*, (1992) found reaction times to be between 140 and 200 ms depending on the stimulus.

However, despite a *plausible* value for "W", it is still an invalid measure of the moment of conscious awareness. The results of this experiment confirmed those of previous research: the subjective report of "W" is imprecise and inaccurate and is a major limiting factor in any iteration of the Libet paradigm (Banks & Isham, 2008; Lau *et al.*, 2004; Lau *et al.*, 2007). This inaccuracy was evidenced by participants reporting "W" to occur after "M" in some instances – confirming their perception of time was poor. The inaccuracy of the reporting of "W" is due to the subjective and post-hoc nature of the report. In order to try remember the moment of conscious awareness, a great deal of concentration is required and is therefore vulnerable to many factors such as fatigue and true understanding of the task. Banks & Isham (2008) and Lau *et al.*, (2007) have already shown how vulnerable this perception of time is, and the inconsistencies, such as reporting "W" to occur after the action "M" in the reports of "W" recorded in this research were thus expected. Despite measures to mitigate fatigue and ensure understanding during data collection, these are not always guaranteed. Fatigue is a strong confounding factor – the setup of the EEG and the experiment itself was very long, increasing the likelihood of fatigue playing a role in the inaccurate reports.

It is recommended that future research move away from this subjective means of measuring conscious intent and rather focus on developing more objective measures of time-locking the moment of conscious awareness. It is for this reason investigating the introduction of eye tracking into the Libet paradigm was a secondary objective.

Eye tracking holds a lot of promise in time-locking the moment of conscious awareness, as previous research has shown the eyes to respond in various ways to changes in cognition (Anantrasirichai *et al.*, 2016; Blignaut, 2009; Einhäuser *et al.*, 2010; Salvucci & Goldberg, 2000; Wierda *et al.*, 2012). Despite the current body of literature showing promise in the eyes providing a gateway to the inner processes in the brain, the results of this experiment did not allow for any inferences to be made. The number of eye events occurring 500 ms before the

action were too low for possible inferences to be made about the correlation between eye movements and the moment of conscious awareness of the decision to move. This poor result is not representative of similar work into the relationship between the eyes and cognition. This poor result can be attributed to there being no correlation at all between the conscious awareness of a decision and a corresponding change in the eyes. However, the confounding factors and subsequent limitations around the eye tracking data collection need to be considered as well. The eye tracker was not calibrated per person, immediately eliminating the possibility of investigating eye fixations (i.e. determining where a person was looking at a particular time) which is an important and useful source of information. The trials were too short to allow for pupil dilation to be included in the analysis. Further, the sampling rate of the eye tracker was 30 Hz below the accepted rate for research. A higher sampling rate means more datapoints are collected per second. The threshold for the identification of a saccade is that it must extend over more than three points. However, with a lower sampling rate, the chance of a saccade being rejected is high. This is due to there being fewer data points per suspected eye event compared to a recording with a higher sampling rate. A higher sampling rate means more data points are recorded per second. This greater resolution is useful when trying to identify events are very short in duration.

Eye blinks were also removed from the continuous as they present in missing data and cannot be included in the analysis. Inferences cannot be made about eye movements when there is no data. The possibility that missing data could coincide with the occurrence of eye movement corresponding to changes in cognition cannot be eliminated. However, no conclusions can be drawn at this point as there were no data on which these assumptions can be further investigated.

Much of the criticism around the Libet paradigm has been addressed. However, despite accurately classifying action using only subconscious EEG data, there is still little understanding of the mechanisms and features the model identified during its training. The "black-box" method is inherent of machine learning methods. However, there are ways of further understanding the CNN model.

Further, it cannot be confirmed whether the subconscious is acting deliberately, or if these features are the results of random neural fluctuations as found in the results of Murakami *et al.* (2014) and Schurger *et al.* (2012). This is difficult to determine on account of the participant being primed to make a binary choice, where a specific decision is inevitable – albeit it deliberate or random. This is also not representative of a real-world scenario. This is a general

flaw in the experiment set up carried over into this research (Khalighinejad *et al.*, 2018). It must be noted that improving the Libet experiment *itself* was not an objective of this research – the goal was to see if machine learning can be used to get more reliable results using the original paradigm. Further research is needed to determine the cause of these subconscious features relating to decision-making. Another avenue of future research is to create real-world scenarios, complete with consequence of any decisions made; such as that of Maoz *et al.* (2017). Another example of a real-world application is that of brain-computer interface systems such as prosthetics or exoskeletons for PLWD.

5.2.1 Implications for rehabilitation: brain-computer interfaces

Despite successfully using machine learning for classifying actions using only subconscious EEG data, it is necessary to consider the application of this result to the field of brain-computer interface (BCI) systems. In classifying decisions ~ one second before conscious awareness, precedence has been set for using this method to predict actions earlier in a BCI set up. In classifying actions earlier, these BCI systems can imitate more realistic real-time thought-action movements, without any kind of delay. (Bai *et al.*, 2011)

BCI systems, including prosthetics are able to bypass the 'normal' anatomical pathways reserved for actions (that is from brain to limb). These normal pathways could be affected centrally (brain) or peripherally (amputations, nerve damage).

There is a need for further research in identifying the specific features used for action classification in both healthy and unhealthy populations; as well as further research to localise 'where' and 'when' the decisions arise. In localising where the decisions arise, guidance can be given regarding the implantation of the chips needed for BCI prosthetics and systems. In localising the exact timing of the features, this can better guide the algorithms in separating intended actions as opposed to ideation about action. Further, this experiment involved simple actions (that is clicking "left" or "right") – relative to the action complexity needed for dexterous arm movements required for normal function. Future experiments should involve more complex movements, in order to extract a coherent array of relevant features in subconscious EEG data in order to support full integration of the recipients into their activities of daily living (ADLs) and into the community.

In the event of peripheral nerve damage (including amputations) – these features will generalisable from the 'normal' populations, as the brain has not been affected. Earlier detection of these features for movement will result in more fluid and natural movements. This functionality is especially important in daily tasks where fast reactions are necessary, such as driving or playing a sport. This functionality is also important in general ADLs such as dressing, and grooming in that these tasks can take up a lot of time – hindering the participants ability to completely engage in their occupation, hobbies and/or community involvement. This limitation in activities and participation is due to much of their time and energy being spent on trying to complete simple tasks. However, in the event of a central lesion – the results will not be generalisable as trauma to the brain has been proven to cause alterations in the EEG (Daly *et al.*, 2014; Konishi *et al.*, 1995; Schomer & Lopes da Silva, 2011).

There are many avenues for furthering and improving on this research. The limitations to this research are outlined in the following section.

5.3 Summary of limitations

Despite successfully using machine learning for classifying actions using only subconscious EEG data, there is still little understanding of the mechanisms and features the model identified during its training. Further research, through the visualisation of the CNN model's learned features, is necessary to understand the exact mechanisms governing the subconscious.

Another limitation is that frames of ~300 - 600 ms of subconscious EEG data were fed into the model to make the classification, resulting in imprecise timing. Visualising the model's learned features is a means of understanding what features the model used, where these features are in the brain, and the exact time they arose.

The experiment is far removed from the real-world, and it can be argued the person is still primed in that they are told to choose between two pre-defined decisions (Khalighinejad et al., 2018). The experiment is also missing the third criterion of a free-willed decision: consequence. Another limitation is the small sample size used in this experiment and others attempting to replicate and improve the original Libet paradigm (Fried *et al.*, 2011; Soon *et al.*, 2008, 2013). This research aimed for 30 participants, however, due to technical issues and time constraints the data collection stopped after 21 participants. Despite these limitations, the average number of participants exceeded that of other studies centered around the Libet paradigm. The range of the number of participants was 5 to 40 participants. The mean number of participants in

these 13 studies was 14.5 (Alexander *et al.*, 2016; Banks & Isham, 2008; Bechara *et al.*, 1997; Fried *et al.*, 2011; Jo *et al.*, 2015, 2014; Libet *et al.*, 1983; Miller *et al.*, 2011; Rigoni *et al.*, 2013; Schultze-Kraft *et al.*, 2016; Soon *et al.*, 2008, 2013; Trevena & Miller, 2010; Verbaarschot *et al.*, 2016).

The biggest limitation of this research however, is the precision of the eye tracking data. The eye tracker was not calibrated for each person as well as the sampling rate of the eye tracker being lower than the accepted sampling rate of 120Hz in eye tracking research. On account of this, we cannot conclusively say where the person was looking, as the internal metrics for the calculations can be off by an indeterminable margin. It is for this reason that the useful fixation data was not included in the analysis for this research. Fortunately, the eye velocity data is less dependent on the calibration accuracy and analysis was able to be made using the eye tracking data. The low sampling rate also meant fewer data points were collected per second, which can increase the probability of a saccade being incorrectly rejected.

Suggestions to address these limitations are made in the next section.

5.4 Future work and recommendations

Future research can be divided into two focus groups: improving the data collection and improving the data analysis.

Improving the data collection methods involves creating a more real-world scenario with inclusion of more complex movements and/or consequence in the experiment such as that of Maoz et al. (2017). The research of Maoz et al. (2017) included a moral decision and a consequential component to it. A more real-world situation could also be achieved by setting up online deep learning processes – for real time predictions. Instead of collecting the data, doing the pre-processing, labelling it and inputting it into the model offline; the predictions could be made before the actions in real-time. This will aid real-time classification-action in BCI systems as well.

Improving the data analysis is two-fold. The immediate next step should be predicting earlier – that is feeding longer windows of the EEG data to classify actions using earlier sections of the subconscious EEG data. Soon et al. (2008) made predictions up to seven seconds before conscious awareness using a fMRI data, which is an excellent benchmark which research into EEG coupled with machine learning should strive towards.

Further, improving the data analysis using machine learning would be to improve interpretability of the machine learning model. This can be achieved through feature extraction. Feature extraction is a method which visualises the 'what', 'where' and 'when' in terms of features in the input the model identifies in order to make its classifications. This feature extraction should also be repeated on populations of people that have suffered from cerebral palsy, epilepsy, a traumatic brain injury or similar (Daly *et al.*, 2014; Schomer & Lopes da Silva, 2011;Konishi *et al.*, 1995). These trauma to the brain can cause changes that would mean the features from a healthy population are not generalisable to other populations. These features would affect the application of the BCI systems to populations that have suffered such injuries.

Localising where in the input the model is finding features can also help reduce the number of electrodes needed to conduct the experiment. Setting up an EEG experiment with 128 electrodes as well as processing all the data is incredibly time consuming. Reducing the number of electrodes needed for the experiment can speed up the EEG data collection and analysis process, allowing more time for interesting analytical methods. Localising these features will also contribute guiding the place of BCI chips to monitor intent around actions as well as contribute more generally to the field neuroscience and decision-making processes.

The addition of eye tracking into the Libet paradigm still holds promise, based on the current literature (Anantrasirichai *et al.*, 2016; Einhäuser *et al.*, 2010; Salvucci & Goldberg, 2000; Wierda *et al.*, 2012). Recommendations for future work include a stringent protocol including calibration of the eye tracker for each participant. A higher sampling rate (greater than 120 Hz) is also advised. Ensuring these two factors are incorporated would allow for more of the components of eye tracking data to be available for analysis. Slight adaptations to the Libet paradigm could be made to ensure each trial is long enough to allow for pupil dilation to be included for consideration. Using eye tracking as a means of separating conscious and subconscious brain activity will contribute greatly to our understanding of the underlying operations of each of these components.

These recommendations are focussed on contributing to the understanding of free will, however, furthering this line of research can also lead to real-world, tangible benefits relating to the lives of people living with disabilities.

5.5 Impact of this research

This research has shown the inclusion of deep learning is an important tool to use in the research revolving around the Libet paradigm, as evidenced by the success of the CNN model. The CNN model has shown that there are interesting features in the subconscious EEG data that relates to our decision-making. This not only contributes to our understanding of the brain as these features will also be useful in furthering research into BCI systems, to improve the lives of PLWD.

Despite the poor results, the body of literature around eye movements and corresponding changes in cognition still supports this has a plausible avenue to pursue in terms of using eye tracking to time-lock the moment of conscious awareness of a decision.

Further research should take place using machine (and deep) learning to further our understanding of the brain's processes and the true nature of free will. Focused investigation into the response of the eyes relative to a decision entering conscious awareness should be made.

6 Conclusion

The following chapter summarises the key findings and implications thereof found in this research.

This research has shown machine learning can be used to develop some understanding about the nature of free will. There is still much to be done in terms of improving the experiment itself, and the methods of analysing the data. However, this research was a proof of concept: to establish whether machine learning can contribute to the current literature. The deep learning neural network, viz. the convolutional neural network employed in this research achieved almost perfect accuracy in classifying an action using subconscious EEG data. This shows that machine learning, with a specific focus on deep learning is an important tool to include in all research in this domain going forward.

Is there free will? This research contributes significantly to the stance that says there is no free will. The results show the decision arises in the subconscious, of which our conscious selves have no control. This proof of features relating to decision-making in the subconscious is important in the context of BCI systems and improving the lives of PLWD. The early detection of action can improve real-time movements and improve functionality.

However, further research is necessary; with a focus on more complex movements and the addition of consequence to future experiments, as well feature visualisation to better understand the features the CNN model has identified during its learning process in order to complete the task. Feature visualisation will improve our understanding of the brain (and the subconsciousness's) underlying mechanisms as well as improve fluidity and speed of the current BCI systems.

Although eye tracking still holds promise in marking the moment of conscious awareness in terms of the current literature, the results of this experiment cannot be considered definitive on account of the limitations of this research. More research is needed focussing on data collection with higher sampling rates and more explorative work into other components of eye tracking data – beyond the focus on saccades as was done in this research. Time-locking the moment of conscious awareness will improve our understanding of the separation of the conscious and subconscious and contribute to the knowledge around decision-making processes.

In conclusion, there is evidence supporting the notion that free will is an illusion. Machine learning is an important tool in various areas of research and in determining the nature of free

will. Further work is needed to understand the correlation between the eyes and the moment of conscious awareness.

7 References

7 best: the human eye images. n.d. [Online], Available: https://clipartxtras.com/%22%3Eclipartxtras.com [2018, November 30].

Aggarwal, V., Acharya, S., Tenore, F., Shin, H.C., Etienne-Cummings, R., Schieber, M.H. & Thakor, N. V. 2008. Asynchronous decoding of dexterous finger movements using M1 neurons. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 16(1):3–14.

Ahirwal, M.K. & Londhe, N.D. 2012. Power Spectrum Analysis of EEG Signals for Estimating Visual Attention. *International Journal of Computer Application*. 42(15):22–25.

Alexander, P., Schlegel, A., Sinnott-Armstrong, W., Roskies, A.L., Wheatley, T. & Tse, P.U. 2016. Readiness potentials driven by non-motoric processes. *Consciousness and Cognition*. 39:38–47.

Allingham, J. 2018.

Amann, B.L., Pogarell, O., Mergl, R., Juckel, G., Grunze, H., Mulert, C. & Hegerl, U. 2003. EEG abnormalities associated with antipsychotics: A comparison of quetiapine, olanzapine, haloperidol and healthy subjects. *Human Psychopharmacology*. 18(8):641–646.

An, Y.W., Lobacz, A.D., Lehmann, T., Baumeister, J., Rose, W.C., Higginson, J.S., Rosen, J. & Swanik, C.B. 2019. Neuroplastic changes in anterior cruciate ligament reconstruction patients from neuromechanical decoupling. *Scandinavian Journal of Medical Science Sports*. 29:251–258.

Anantrasirichai, N.., Gilchrist, I.D.. & Bull, D.R. 2016. Fixation identification for low-sample-rate mobile eye trackers. In IEEE *2016 IEEE International Conference on Image Processing (ICIP)*. 3126–3130.

Anders, P., Lehmann, T., Müller, H., Grønvik, K.B., Skjæret-maroni, N., Baumeister, J. &

Vereijken, B. 2018. Exergames Inherently Contain Cognitive Elements as Indicated by Cortical Processing. 12(May):1–8.

Ang, K.K., Sui, K., Chua, G., Phua, K.S., Wang, C., Chin, Z.Y., Wee, C., Kuah, K., et al. 2015. A Randomized Controlled Trial of EEG-Based Motor Imagery Brain-Computer Interface Robotic Rehabilitation for Stroke.

Arnesano, C. 2009. *Intraoperative Detection of Awareness*. [Online], Available: http://bme240.eng.uci.edu/students/09s/arnesanc/bispectral_anslysis.html [2018, October 06].

Bai, O., Rathi, V., Lin, P., Huang, D., Battapady, H., Fei, D.Y., Schneider, L., Houdayer, E., et al. 2011. Prediction of human voluntary movement before it occurs. *Clinical Neurophysiology*. 122(2):364–372.

Bandarabadi, M., Teixeira, C.A., Rasekhi, J. & Dourado, A. 2015. Clinical Neurophysiology Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology*. 126(2):237–248.

Banks, W.P. & Isham, E.A. 2008. We infer rather than perceive the moment we decided to act. *Psychological Science*. 20(1):17–22.

Banoczi, W. 2005. How some drugs affect the electroencephalogram (EEG). *American journal of electroneurodiagnostic technology*. 45(2):118–129.

Barry, R.J., Clarke, A.R., Johnstone, S.J. & Rushby, J.A. 2008. Timing of caffeine's impact on autonomic and central nervous system measures: Clarification of arousal effects. *Biological Psychology*. 77(3):304–316.

Bechara, A., Damasio, H., Tranel, D., Damasio, A.R., Bushnell, M.C., Matthews, P.M. & Rawlins, J.N.P. 1997. Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*. 275(5304):1293–1295.

van Biljon, E. 2018.

Blignaut, P. 2009. Fixation identification : The optimum threshold for a dispersion algorithm. *Attention, Perception & Pscychophysics*. 71(4):881–895.

Bode, S., Murawski, C., Soon, C.S., Bode, P., Stahl, J. & Smith, P.L. 2014. Demystifying "free will": The role of contextual information and evidence accumulation for predictive brain activity. *Neuroscience and Biobehavioral Reviews*. 47:636–645.

Brass, M. & Haggard, P. 2008. The what, when, whether model of intentional action. *Neuroscientist*. 14(4):319–325.

Broglio, S.P., Moore, R.D. & Hillman, C.H. 2011. A history of sport-related concussion on event-related brain potential correlates of cognition. *International Journal of Psychophysiology*. 82:16–23.

Burns, K. & Bechara, A. 2007. Decision Making and Free Will: A Neuroscience Perspective. *Behavioral Sciences & the Law*. 25:263–280.

Caffeine. n.d. [Online], Available: https://www.drugs.com/ingredient/caffeine.html [2018, October 17].

Chella, F., Pizzella, V., Zappasodi, F. & Marzetti, L. 2016. Impact of the reference choice on scalp EEG connectivity estimation. *Journal of Neural Engineering*. 13:1–34.

Chollet, F. 2018. *Keras: The Python Deep Learning library*. [Online], Available: https://keras.io/ [2019, August 01].

Clozapine. 2018. [Online], Available: https://www.drugs.com/clozapine.html [2018, October 14].

Cohen, M.X. 2014. *Analysing neural time series data: theory and practise*. 1st ed. London: The MIT Press.

CS1114 Section 6: Convolution. 2013. [Online], Available: https://www.cs.cornell.edu/courses/cs1114/2013sp/sections/S06_convolution.pdf [2018, December 08].

Daly, I., Faller, J., Scherer, R., Sweeney-Reed, C.M., Nasuto, S.J., Billinger, M. & Müller-Putz, G.R. 2014. Exploration of the neural correlates of cerebral palsy for sensorimotor BCI control. *Frontiers in Neuroengineering*. 7(July):1–11.

Deep Learning Indaba. 2018a. *Practical 1: Deep Feed-forward Networks*. Github.

Deep Learning Indaba. 2018b. *Practical 2: Convolutional Networks*. Github. [Online], Available: https://github.com/deep-learning-indaba/indaba-2018/Practical_2_Convolutional_Neural_Networks.ipynb [2018, December 07].

Delmonte, D.W. & Kim, T. 2011. Anatomy and physiology of the cornea. *Journal of Cartaract & Refractive Surgery*. 37:588–598.

Delorme, A. & Makeig, S. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 134(1):9–21.

Delorme, A. & Makeig, S. 2018. *EEGLAB Tutorial:Performing Independent Component Analysis of EEG data*. [Online], Available: https://sccn.ucsd.edu/wiki/Chapter_09:_Decomposing_Data_Using_ICA [2019, October 05].

Delorme, A. & Makerig, S. n.d. *EEGLAB Wiki*. [Online], Available: https://sccn.ucsd.edu/wiki/EEGLAB_Wiki [2018, August 01].

Delorme, A., Palmer, J., Onton, J., Oostenveld, R. & Makeig, S. 2012. Independent EEG sources are dipolar. *PLoS ONE*. 7(2).

Dias, A.M. & Lavazza, A. 2016. Free will and neuroscience: from explaining freedom away to new ways of operationalizing and measuring it. *Frontiers in Human Neuroscience*. 10(October):1–17.

Dien, J. 1998. Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods, Instruments, and Computers*. 30(1):34–43.

Doyle, R. n.d. *The Information Philosopher*. [Online], Available: http://www.informationphilosopher.com/freedom/libet_experiments.html [2018, July 04].

Duchowski, A.T. 2007. *Eye Tracking Methodology*. 2nd ed. London: Springer-Verlag.

Eagleman, D.M. 2004. The where and when of intention. *Science*. 303:1144–1147.

Eagleman, D.M. 2011. *Incognito: the secret lives of the brain*. 1st ed. New York: Vintage Books.

Ehlers, C.L., Wall, T.L. & Schuckit, M.A. 1989. EEG spectral characteristics following ethanol administration in young men. *Electroencephalography and Clinical Neurophysiology*. 73(3):179–187.

Einhäuser, W., Koch, C. & Carter, O.L. 2010. Pupil dilation betrays the timing of decisions. *Frontiers in Human Neuroscience*. 4(18):1–9.

Fergus, P., Hignett, D., Hussain, A., Al-Jumeily, D. & Abdel-Aziz, K. 2015. Automatic epileptic seizure detection using scalp EEG and advanced artificial intelligence techniques. *BioMed Research International*. 2015.

Fergus, P., Hussain, A., Hignett, D., Al-Jumeily, D., Abdel-Aziz, K. & Hamdan, H. 2016. A machine learning system for automated whole-brain seizure detection. In *Applied Computing and Informatics*. 70–89.

Fookes, C. 2018a. *Selective serotonin reuptake inhibitors*. [Online], Available: https://www.drugs.com/drug-class/ssri-antidepressants.html.

Fookes, C. 2018b. *Serotonin-norepinephrine reuptake inhibitors*. [Online], Available: https://www.drugs.com/drug-class/ssnri-antidepressants.html [2018, October 14].

Fookes, C. 2018c. *Tricyclic Antidepressants*. [Online], Available: https://www.drugs.com/drug-class/tricyclic-antidepressants.html [2018, October 14].

Fookes, C. 2018d. *Benzodiazepines*. [Online], Available: https://www.drugs.com/drug-class/benzodiazepines.html [2018, October 14].

Fried, I., Mukamel, R. & Kreiman, G. 2011. Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron*. 69(3):548–562.

Fukushima, K. 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*. 36(4):193–202.

Gabard-Durnam, L.J., Leal, A.S.M., Wilkinson, C.L. & Levin, A.R. 2018. The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data. *Frontiers in Neuroscience*. 12(97):1–24.

Géron, A. 2017. *Hands-On Machine Learning with Scikit-Learn & TensorFlow: Concepts, Tools and Techniques to Build Intelligent Systems*. 1st ed. ed. N. Tache (ed.). Sebastopol: O'Reilly Media, Inc.

Gomes, G. 2007. Free will, the self and the brain. *Behavioral Sciences & the Law*. 25(1):221–234.

Goodfellow, I., Bengio, Y. & Courville, A. 2016. *Deep Learning*. MIT Press. [Online], Available: http://www.deeplearningbook.org.

Goodman, C. 2002. Technique for the Analysis of Evoked and Background EEG Activity applied to Young and Elderly Subjects. Hebrew Univeristy.

Haggard, P. 2011. Decision time for free will. *Neuron*. 69(3):404–406.

Harris, S. 2012. *Free Will*. 1st ed. New York: Free Press.

Hawking, S.W. 1993. *Black Holes and Baby Universes and Other Essays*. New York: Bantam Books.

Heming, R.I., Glover, B.J., Koeppl, B., Phillips, R.L. & London, E.D. 1994. Cocaine-induced increases in EEG alpha and beta activity: Evidence for reduced cortical processing. *Neuropsychopharmacology*. 11(1):1–9.

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H. & Weijer, J. Van De. 2011. *Eye Tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Holsheimer, J. & Feenstra, B.W. 1977. Volume conduction and EEG measurements within the brain: a quantitative approach to the influence of electrical spread on the linear relationship of activity measured at different locations. *Electroencephalography and Clinical Neurophysiology*. 43(1):52–58.

How to Read an EEG. n.d. [Online], Available: https://www.epilepsy.com/learn/diagnosis/eeg/how-read-eeg [2019, March 22].

Hubel, D.H. 1959. Single unit activity in striate cortex of unrestrained cats. *The Journal of Physiology*. 147:226–238.

Hubel, D.H. & Wiesel, T.N. 1959. Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*. 148:574–591.

Hubel, D.H. & Wiesel, T.N. 1968. Receptive Fields and Functional Architecture of Monkey Striate COrtex. *The Journal of Physiology*. 195:215–243. [Online], Available: http://www.ncbi.nlm.nih.gov/pubmed/19525561.

Huettel, S.A., Song, A.W. & McCarthy, G. 2004. *Functional Magnetic Resonance Imaging*. Second ed. Sunderland: Sinauer Associates, Inc.

Huettel, S.A., McKeown, M.J., Song, A.W., Hart, S., Spencer, D.D., Allison, T. & McCarthy, G. 2004. Linking Hemodynamic and Electrophysiological Measures of Brain Activity: Evidence from Functional MRI and Intracranial Field Potentials. *Cerebral Cortex*. 14(2):165–173.

iMotions. 2018.

James, W. 1981. *The Principles of Psychology*. Cambridge, MA: Harvard University Press.

Jo, H.G., Wittmann, M., Borghardt, T.L., Hinterberger, T. & Schmidt, S. 2014. First-person approaches in neuroscience of consciousness: Brain dynamics correlate with the intention to act. *Consciousness and Cognition*. 26(1):105–116.

Jo, H.G., Hinterberger, T., Wittmann, M. & Schmidt, S. 2015. Do meditators have higher awareness of their intentions to act? *Cortex*. 65:149–158.

Johannesen, J.K., Bi, J., Jiang, R., Kenney, J.G. & Chen, C.-M.A. 2016. Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. *Neuropsychiatric Electrophysiology*. 2(1):3.

Juhola, M. 1991. Median filtering is appropriate to signals of saccadic eye movements. *Computers in Biology and Medicine*. 21(1–2):43–49.

Karpathy, A. n.d. *Convolutional Neural Networks for Visual Recognition*. [Online], Available: http://cs231n.github.io/neural-networks-1/ [2018, October 29].

Kemp, B.J. 1973. Reaction time of young and elderly subjects in relation to perceptual deprivation and signal-on versus signal-off conditions. *Developmental Psychology*. 8(2):268–272.

Khalighinejad, N., Schurger, A., Desantis, A., Zmigrod, L. & Haggard, P. 2018. Precursor processes of human self-initiated action. *NeuroImage*. 165(September 2017):35–47.

Koch, C. 2012. *Consciousness: Confessions of a Romantic Reductionist*. First ed. MIT Press. [Online], Available: http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=12877.

Konishi, T., Naganuma, Y., Hongou, K., Murakami, M., Yamatani, M. & Okada, T. 1995. Effects of Antiepileptic Drugs on EEG Background Activity in Children with Epilepsy: Initial Phase of Therapy. *Clinical EEG and Neuroscience*. 26(2):113–119.

Kornhuber, H. & Deecke, L. 1965. 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.

Krieghoff, V., Waszak, F., Prinz, W. & Brass, M. 2011. Neural and behavioral correlates of intentional actions. *Neuropsychologia*. 49(5):767–776.

Kuhn, R.L. & Koch, C. 2014. [Online], Available: https://www.youtube.com/watch?v=PZ-DPsy0eaw.

Lansbergen, M.M., Dumont, G.J.H., Van Gerven, J.M.A., Buitelaar, J.K. & Verkes, R.J. 2011. Acute effects of MDMA (3,4-methylenedioxymethamphetamine) on EEG oscillations: Alone and in combination with ethanol or THC (delta-9- tetrahydrocannabinol). *Psychopharmacology*. 213(4):745–756.

Laplace, P.S. 1902. *A philosophical essay on probabilities*. Translated ed. New York: John Wiley and Sons.

Lau, H.C., Rogers, R.D., Haggard, P. & Passingham, R.E. 2004. Attention to Intention. *Science*. 303(5661):1208–1210.

Lau, H.C., Rogers, R.D. & Passingham, R.E. 2007. Manipulating the Experienced Onset of Intention after Action Execution. *Journal of Cognitive Neuroscience*. 19(1):81–90.

Libet, B. 1999. Do we have free will? *Journal of Consciousness Studies*. 6(8):47–57.

Libet, B., Gleason, C.A., Wright, E.W. & Pearl, D.K. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain*. 106(3):623–642.

Logesparan, L., Casson, A.J. & Rodriguez-Villegas, E. 2012. Optimal features for online seizure detection. *Medical and Biological Engineering and Computing*. 50(7):659–669.

Logesparan, L., Casson, A.J., Imtiaz, S.A. & Rodriguez-Villegas, E. 2013. Discriminating between best performing features for seizure detection and data selection. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology*

Society, EMBS. 1692–1695.

Mahajan, S.. 2018. *Does Gradient Descent Algo always converge to the global minimum?* [Online], Available: https://www.quora.com/Does-Gradient-Descent-Algo-always-converge-to-the-global-minimum [2018, November 20].

Makeig, S. & Onton, J. 2011. ERP Features and EEG Dynamics: An ICA perspective. In E.S. Kappenman & S.J. Luck (eds.). New York: Oxford University Press *The Oxford Handbook of Event-Related Potential Components*.

Makeig, S., Bell, A.J., Jung, T.-P. & Sejnowski, T.J. 1996. Independent Component Analysis of Electroencephalographic Data. 145–151.

Maoz, U., Yaffe, G., Koch, C. & Mudrik, L. 2017. *Neural precursors of decisions that matter — an ERP study of deliberate and arbitrary choice*.

Meng, J., Mundahl, J.H., Streitz, T.D., Maile, K., Gulachek, N.S., He, J. & He, B. 2017. Effects of Soft Drinks on Resting State EEG and Brain-Computer Interface Performance. *IEEE Access*. 5:18756–18764.

Miller, J., Shepherdson, P. & Trevena, J. 2011. Effects of clock monitoring on electroencephalographic activity: Is unconscious movement initiation an artifact of the clock? *Psychological Science*. 22(1):103–109.

Mirowski, P., Madhavan, D., Lecun, Y. & Kuzniecky, R. 2009. Clinical Neurophysiology Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology*. 120(11):1927–1940.

Mitchell, T. 1998. *Machine Learning*. New York: McGraw-Hill.

Multum, C. 2018. *Epinephrine injection*. [Online], Available: https://www.drugs.com/mtm/epinephrine-injection.html [2018, December 05].

Murakami, M., Vicente, M.I., Costa, G.M. & Mainen, Z.F. 2014. Neural antecedents of self-initiated actions in secondary motor cortex. *Nature Neuroscience*. 17(11):1574–1582.

Ng, A. n.d. *Machine Learning*. Stanford University on Coursera. [Online], Available: https://www.coursera.org/learn/machine-learning [2018, October 27].

Nurse, E.S., Karoly, P.J., Grayden, D.B. & Freestone, D.R. 2015. A Generalizable Brain-Computer Interface (BCI) Using Machine Learning for Feature Discovery. *PLoS ONE*. 10(6):1–23.

Olsen, A. 2012. *The Tobii I-VT Fixation Filter Algorithm description*.

Oxford South African Concise Dictionary. 2010. 2nd ed. Cape Town: Oxford University Press.

Peirce, J.W. 2007. PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*. 162:8–13.

Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E. & Lindeløv, J.K. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*. 51:195–203.

Perlovsky, L. 2011. Consciousness and Free Will, A Scientific Possibility Due to Advances in Cognitive Science. *WebmedCentral Psychology*. 2(2):1–9. [Online], Available: http://www.webmedcentral.com/article_view/1539.

Pham, V., Bluche, T., Kermorvant, C. & Louradour, J. 2014. Dropout improves Recurrent Neural Networks for Handwriting Recognition.

Powers, K.C., Kalmar, J.M. & Cinelli, M.E. 2014. Dynamic stability and steering control following a sport-induced concussion. *Gait and Posture*. 39:728–732.

Rigoni, David. Brass, Marcel. Roger, Clémence. Vidal, Franck. Sartori, G. 2013. Top-down modulation of brain activity underlying intentional action and its relationship with awareness of intention: an ERP/ Laplacian analysis. *Experimental Brain Research*. 229:347–357.

Rothman, M. 2015. *"Back to the Future" 30th Anniversary: Neil deGrasse Tyson Talks Time Travel*. [Online], Available: https://abcnews.go.com/Entertainment/back-future-30th-

anniversary-neil-degrasse-tyson-talks/story?id=32191481 [2019, November 11].

Salinsky, M.C., Oken, B.S. & Morehead, L. 1995. Intraindividual analysis of antiepileptic drug effects on EEG background rhythms. *Clinical Neurophysiology*. 90(3):186–193.

Salvucci, D.D. & Goldberg, J.H. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. 71–78.

Samuel, A.L. 1959. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*. 3(3):211–229.

Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., et al. 2017. Deep Learning With Convolutional Neural Networks for EEG Decoding and Visualization. *Human Brain Mapping*. 38:5391–5420.

Schmidt, H. n.d. *Nervous System - Neuron: Nerve Cell*. [Online], Available: https://myclass.theinspiredinstructor.com/science/health_diagrams/Neuron_Label.htm [2018, March 10].

Schmidt, S., Jo, H.G., Wittmann, M. & Hinterberger, T. 2016. 'Catching the waves' – slow cortical potentials as moderator of voluntary action. *Neuroscience and Biobehavioral Reviews*. 68:639–650.

Schomer, D.L. & Lopes da Silva, F.H. Eds. 2011. *Niedermeyer's Electroencephalography: Basic principles , clinical applications and related fields*. 6th ed. Philadelphia: Lippincott Williams & Wilkins.

Schultze-Kraft, M., Birman, D., Rusconi, M., Allefeld, C., Görgen, K., Dähne, S., Blankertz, B. & Haynes, J.-D. 2016. The point of no return in vetoing self-initiated movements. *Proceedings of the National Academy of Sciences*. 113(4):1080–1085.

Schurger, A., Sitt, J.D. & Dehaene, S. 2012. An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*. 109(42):E2904–E2913.

Schurger, A., Mylopoulos, M. & Rosenthal, D. 2016. Neural antecedents of spontaneous voluntary movement: a new perspective. *Trends in Cognitive Sciences*. 20(2):77–79.

Sieppman, M. & Kirch, W. 2002. Effect of Caffeine on Topographic Quantitative EEG. *Neuropsychobiology*. 45:161–166.

Smith, E.. n.d. *Introduction to EEG*. [Online], Available: https://www.ebme.co.uk/articles/clinical-engineering/56-introduction-to-eeg [2018, July 30].

Soon, C.S., Brass, M., Heinze, H.J. & Haynes, J.D. 2008. Unconscious deteminants of free decisions in the human brain. *Nature Neuroscience*. 11(0):543–545.

Soon, C.S., He, A.H., Bode, S. & Haynes, J.-D. 2013. Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences*. 110(15):6217–6222.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*. 15:1929–1958.

Teel, E.F., Ray, W.J., Geronimo, A.M. & Slobounov, S.M. 2014. Residual alterations of brain electrical activity in clinically asymptomatic concussed individuals: An EEG study. *Clinical Neurophysiology*. 125:703–707.

Template 2D layouts for plotting. 2018. [Online], Available: http://www.fieldtriptoolbox.org/template/layout [2018, October 01].

Teplan, M. 2002. Fundamentals of EEG measurement. *Measurement Science Review*. 2(2):1–11.

Thomson, P.D., Colebatch, J.G., Brown, P., Rothwell, J.C., Day, B.L., Obeso, J.A. & Marsden, C.D. 1992. Voluntary stimulus-sensitive jerks and jumps mimicking myoclonus or pathological startle syndromes. *Movement Disorders*. 7(3):257–262.

Tobii Pro. n.d. *How do Tobii Eye Trackers work?* [Online], Available:

https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-do-tobii-eye-trackers-work/ [2018, November 27].

Trevena, J. & Miller, J. 2010. Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and Cognition*. 19(1):447–456.

Vargas, M. 2004. Libertarianism and the Skepticism about Free Will : Some Arguments against Both. *Philosophical Topics*. 32(1–2):403–426.

Verbaarschot, C., Farquhar, J. & Haselager, P. 2015. Lost in time...The search for intentions and Readiness Potentials. *Consciousness and Cognition*. 33:300–315.

Verbaarschot, C., Haselager, P. & Farquhar, J. 2016. Detecting traces of consciousness in the process of intending to act. *Experimental Brain Research*. 234:1945–1956.

Vinding, M.C., Pedersen, M.N. & Overgaard, M. 2013. Unravelling intention : Distal intentions increase the subjective sense of agency. *Consciousness and Cognition*. 22:810–815.

Volschenk, A.D. 2017. Application of Machine Learning with Electroencephalography in Seizure Detection. Stellenbosch University.

Wierda, S.M., van Rijn, H., Taatgen, N.A. & Martens, S. 2012. Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*. 109(22):8456–8460.

WikiAudio. 2016. [Online], Available: https://www.youtube.com/watch?v=leHYhS22b5g.

Wolpe, N. & Rowe, J.B. 2014. Beyond the “urge to move”: objective measures for the study of agency in the post-Libet era. *Frontiers in Human Neuroscience*. 8(June):1–13.

Yang, J. 2017. *ReLU and Softmax Activation Functions*. [Online], Available: https://github.com/Kulbear/deep-learning-nano-foundation/wiki/ReLU-and-Softmax-Activation-Functions [2018, October 15].

Yang, L., Leung, H., Plank, M., Snider, J. & Poizner, H. 2015. EEG activity during movement planning encodes upcoming peak speed and acceleration and improves the accuracy in predicting hand kinematics. *IEEE Journal of Biomedical and Health Informatics*. 19(1):22–28.

Yang, S., Flores, B., Magal, R., Harris, K., Gross, J., Ewbank, A., Davenport, S., Ormachea, P., et al. 2017. Diagnostic accuracy of tablet-based software for the detection of concussion. *PLoS ONE*. 1–14.

Young, L.R. & Sheena, D. 1975. Survey of eye movement recording methods. *Behaviour Research Methods and Instrumentation*. 7(5):397–429.

Zakeri, Z. 2016. Optimised Use of Independent Component Analysis for EEG Signal Processing. University of Birmingham.

Zeiler, M.D. & Fergus, R. 2013. Visualizing and Understanding Convolutional Networks.

Zhang, L., Cichocki, A. & Amari, S.I. 2004. Self-adaptive blind source separation based on activation functions adaptation. *IEEE Transactions on Neural Networks*. 15(2):1–12.

Zhou, Y. & Chellappa, R. 1988. Computation of optical flow using a neural network. In *IEEE International Conference on Neural Networks*. 71–78.

Appendix A The eyes

A.1 The anatomy of the eye

To aid the understanding of visual attention, an overview of the anatomy of the eye can be useful to understand these movements. The figure below is a basic representation of the human eye.

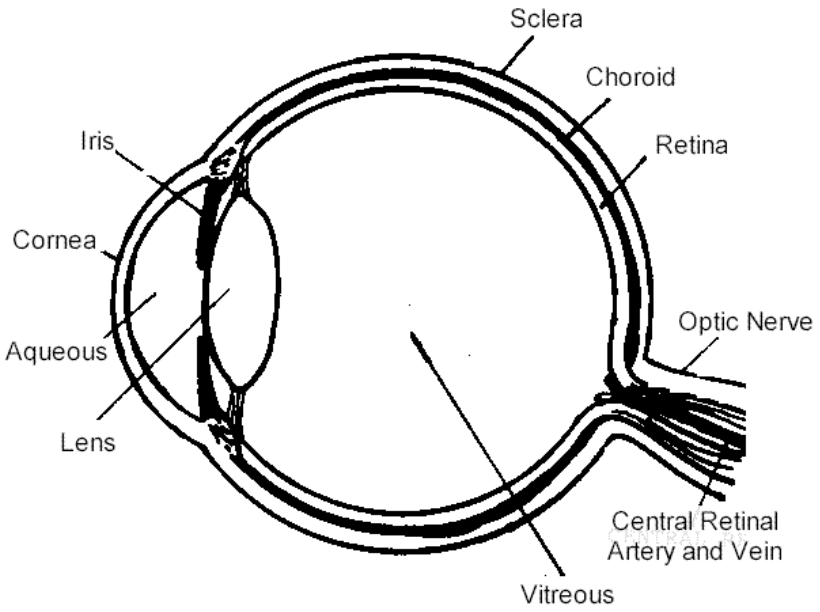


Diagram of the human eye ("7 best: the human eye images", n.d.)

The most important components of the eye, in the field of eye tracking are the retina (and the fovea – not depicted; but falls anatomically close to the marker for the retina in Figure 1), the cornea and the pupil. Each of these components will be discussed in detail. Their functions are summarised as follows in the following table:

The components of the eye relevant to eye tracking systems

Component	Description
Retina	Interior surface on the posterior aspect of the eye. Contains photoreceptors – sensitive to light – which convert light energy to neural impulses, initiating the process of visual perception. (Duchowski, 2007)
Fovea	Point in the retina of high visual acuity for image processing such as required for reading and driving (Duchowski, 2007; Young & Sheena, 1975).
Pupil	Centre of the iris – absorbs light, making it appear darker than the surrounding iris. Provides the avenue through which light can reach the retina. (Young & Sheena, 1975)
Cornea	Avascular, transparent connective tissue that is the eye's primary barrier against infection and irritants (Delmonte & Kim, 2011).

Another important factor to understand when implementing eye tracking systems is the movement of the human eye (Young & Sheena, 1975). These movements are brought by a neural feedback system known as the oculomotor plant, with three categories of control regarding eye movements: the so-called voluntary control, by the occipital cortex; the involuntary control, by the superior colliculus; and the reflexive circuit responsible for maintaining balance signalled by the semi-circular canals. According to Duchowski (2007), the latter of these categories is not relevant to this research and won't be addressed in detail.

There are six muscles responsible for producing eye ball movement (having received efferent input from the above-mentioned brain regions. The table below summarises these muscles and their (basic) functions.

Muscles responsible for producing movement in the eye

Muscle	Function
Medial and lateral recti	Horizontal movements
Superior and inferior recti	Vertical movements
Superior and inferior obliquus	Torsion movements

(Duchowski, 2007)

A.2 The eye movements

Movement of the eye can be refined into two classes: positional and non-positional movements. The basic eye movements above are combined in order to produce the fine movements described below.

Positional movements are as follows:

- Saccadic:
    - These can be voluntary or reflexive and are characterised by rapid eye movements that occur when the gaze moves from one object of focus to another
    - These movements have a sharp acceleration and deceleration, which is up to 40 000 degrees.second-^{2}
- (Smooth) pursuit / slow tracking:
    - This refers to the involuntary "tracking" of a moving object. Assuming the object is moving at an appropriately slow speed, the eye is capable of "matching velocity" to be able to follow the moving object's path
- Fixations:
    - Miniature eye movements (< 1 degree in amplitude) that are present when attempting to stabilise the eye on a stationary object. Fixations are further divided into three categories: drift, tremors and microsaccades. As these terms suggest, the eye is not completely still, but rather experiencing automatic (minuscule) fluctuations in attempting to maintain a steady gaze
- Vergence:
    - The focussing of both eyes over a distance (depth perception).

- Vestibular movements:
    - Automatic, reflexive eye movements compensating for passive or active movements of the head and trunk to ensure a stable “retinal image” is produced, despite the motion.

(Duchowski, 2007; Young & Sheena, 1975)

Only the first these (saccades) are relevant to eye tracking systems. The following non-positional eye movements are also relevant to eye tracking systems.

Non-positional movements are as follows:

- Adaptation:
    - Pupil dilation in response to change in the degree of light entering the eye. This is the only non-positional movement measurable by applicable eye tracking systems
- Accommodation:
    - Lens adjustments to maintain focus on objects of varying distances

(Duchowski, 2007)

Appendix B Machine (and deep) learning

B.1 Features used for support vector machines (SVMs)

The following table lists the common features for seizure detection in EEG data using SVMs.

Common features for seizure detection in EEG data using SVMs

Feature	Description
Peak frequency	Frequency of the highest peak of the Power Spectral Density (PSD).
Median frequency	Separate frequencies in $\delta, \theta, \alpha, \beta$ bands.
Root Mean Square (RMS)	Measure of a signal strength in a given frequency band (amplitude related) $RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N-1} x(i)^2}$ Equation 1 N = no. of samples in the physiological signal x x = EEG of a given channel in a given frequency band.
Sample entropy	Measure of the complexity (amount of information present) of a signal.
Signal energy	Energy distribution per frequency band.

Information referenced from: (Fergus *et al.*, 2015, 2016; Logesparan *et al.*, 2012; Volschenk, 2017)

These pre-determined features are examples of a kind of transformation (i.e. a changed form of the pure signal) of the EEG data, something this research hopes to avoid, by allowing the ANN to determine its own set of features. SVMs have had success in the task of differentiating between two kinds signals (ictal and non-ictal) using these pre-determined features. However, the task of this research is to make a prediction of how one phase will present given a previous phase (i.e. what will happen in the conscious phase, given only the pre-conscious signals). Essentially this task is a form of classification (using pre-conscious EEG data to predict a movement "left"/ "right"), and theoretically a classification algorithm such as SVMs could be used. However, the task of this research is *more* similar to the work being done in seizure *prediction* using feedforward neural networks (Mirowski *et al.*, 2009) such as that already described, recurrent neural networks (Mirowski *et al.*, 2009) and convolutional neural networks (Bandarabadi *et al.*, 2015). The ANNs are used to find neural precursors to the seizure onset (i.e. using the pre-ictal phase to make predictions about the ictal phase – i.e. seizure onset). These are more appropriate to the context of this research and will be investigated further, with the inclusion of a proposed ANN.

B.2 Cost functions

Before explaining the cost function, used interchangeably with loss function, the term 'parameters' (denoted by θ) needs to be clarified. Parameters (not to be confused with hyperparameters which are selected by the programmer) are what the model learns from the data and uses to improve its function (i.e. the hypothesis $h_{(\theta)}$) in order to obtain the best fit of the data; to ensure an appropriate estimated output (the model's calculation of what y could be). The ML model applies a function (the form of which we choose) that takes input training examples x and parameters θ as input. The cost function then assigns a measure of how well the parameters are suited to the task. Through this we can compare how well different parameter configurations perform and we can then select the optimal one (van Biljon, 2018).

In other words, the aim is to learn (i.e. find) the values of θ so that the hypothesis (outputted by the algorithm) best fits the data so that it is able to use its hypothesis to complete the task on a new, unseen example x with as low an error as possible; e.g. classify an unseen x into one of two classes. This will be achieved through the ML model finding the best parameters to obtain the minimum error (cost/loss, denoted by $J(\theta)$). This process of adjusting the parameters to obtain a better hypothesis is done by the ML algorithm on its own. How does it know when it has achieved the optimal parameters to fit the model to the data? Simple, having

(for example) randomly initialised values for θ, it determines the error. Then, it updates its parameters and tries again. This repeats at each step: before updating its parameters: it measures error. This error is the difference between the hypothesis ($h_{(\theta)}$) and the actual target value (i.e. the correct target "answer", viz. y). There is more than one way to approach the calculation of the cost function $\big(J(\theta)\big)$, but for the basis of this explanation, the example of mean squared error (MSE) will be used; which calculates the squared difference between the hypothesis's estimate and the target y (i.e. the y given as an 'answer' in the context of supervised learning). The formula for MSE is as follows, where m is the number of training examples provided:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

Equation 2

B.3 Activation functions

Saturating, non-linear functions:

- Sigmoid (logistic) activation function:
    - The sigmoid activation function scales the values so they lie between the range [0;1]. It is generally used in a classification task with two classes and outputs a probability
    - In general, large negative numbers are scaled towards 0, and large positive numbers scale to 1 (Géron, 2017; Yang, 2017)
    - This tends to be within the neural network. Please see "*Softmax activation function"* for further distinctions between these similar units
    - Saturating functions are such that the activation functions outputs values that are close to the asymptotes of the bounded activation functions (Goodfellow *et al.*, 2016). This is where vanishing or exploding gradients come in, as values <1 multiplied numerous times cause the numbers to become smaller and smaller until they're virtually zero. If each number in the dataset has a value >1, the products will converge to infinity (i.e. "explode"), with the reverse happening when the value is close to one
    - The equation is as follows:

$$g(z) = \frac{1}{1 + e^{-z}}$$

Equation 3

- Softmax activation function:
  - The softmax activation function also scales the values so they lie between the range [0;1]
  - This is generally an output activation unit that has an input vector of more than two elements. The softmax function divides each output in the vector so the total sum of the outputs is equal to one. (Yang, 2017)

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_j}}$$

Where z = vector of inputs to the output, indexed by j

Equation 4

  - The main difference, in terms of the ANN design, between the sigmoid and the softmax activation functions is the position in the ANN's architecture; viz. sigmoid tends to be in the hidden layers and softmax in the output layer. Another key difference is that sigmoid activation functions are applied element wise to the input. Each value inside the input tensor will only affect their corresponding value in the output tensor. The softmax is not applied element wise (van Biljon, 2018)
  - The sigmoid is able to scale over one variable, while the softmax function scales over more than one variable

Non-saturating, non-linear functions:

- Rectified Linear Units (ReLU):
  - This is generally a unit part of the hidden layers

$$f(z) = \max(x, 0)$$

Equation 5

- The ReLU activation function is piece-wise linear – each defined region is linear, bit overall (globally) it is non-linear. It's a simpler activation unit in the following conditions are met:
  - Input < 0 → outputs zero
  - Input > 0 → raw output ∴ input = output. (Yang, 2017)
- This unit has the advantage of faster training, but these units can "die" during training if the learning rate is not appropriate. A large gradient can cause the weights to update during backpropagation in such a way that the neuron is no longer active and any gradients flowing through will automatically be zero. This is irreversible and can happen to as much as 40% of the ANN. (Karpathy, n.d.)

Appendix C Methodology

C.1 Screening tool for volunteers

The following letter was sent to individuals expressing interest in volunteering as participants:

To whom it may concern

You are invited to participate in research projects centered around electroencephalography (EEG) recordings. The aim of this study is to investigate the decision-making processes and the associated brain mechanisms. These brain mechanisms are investigated through the non-invasive recording of the brain's electrical signals with EEG.

The testing process is simple and the experiment set up includes the placing of a 128-electrode cap on your head and then gelling thereof with a blunted needle. The procedure involves no risks; however, you are free to refuse to participate at any point, even after you have signed the consent form. You will be instructed on a simple task to perform during the test. There are two separate tests.

One of the experiments involves the use of an eye tracker which measures changes in pupil dilation as well as pupil movements.

The experiment will take place at the Neuromechanics Unit at Coetzenburg (behind the swimming pool to the right of the gym). Set up and testing will take about 3 hours. Snacks will be provided – please advise on any allergens and/or dietary requirements.

Before volunteering, please ensure you are able to comply with the following conditions. These are in place to try reduce the potential confounding factors that may alter your EEG signals making analysis and comparison difficult.

1. Normal, or corrected to normal vision (contacts only as glasses may cause added reflections which have the potential to confuse the eye tracking system).
2. No gel / hair products / wet hair on the day of testing. Please do not use conditioner within 12 hours of testing.
3. Abstinence from alcohol and stimulants 24 hours prior to testing.
4. No caffeine 2 hours prior to testing (including but not limited to coffee, tea, caffeinated soft drinks).
5. You do not need to disclose your medication history, but if you are taking any of these please decline to participate:

Clozapine

Haloperidol

Tricyclic antidepressant

Benzodiazepines

Antihistamines

Opiate-based analgesics (e.g. Demerol)

Warfarin

Antihypertensive medication

Active chemotherapy or radiation

Any recreational drug use (including but not limited to marijuana)

6. You do not need to disclose your medical history, but if you have a history of any of the following conditions, please decline to participate:

   Epilepsy

   Cerebral palsy

   Concussion(s) within the last 12 months

   Active CNS inflammation

   Previous traumatic brain injury(-ies) and/or tumour(s) [both present and as well as previously removed]

Should you have any questions please don't hesitate to contact Siobhan Hall: +27 (0) 760421618. Alternatively, you can contact the Professor in charge of this research: Prof van den Heever: +27 (0) 21 808 4856.

Thank you for your consideration to participate in this study.

Yours faithfully
Siobhan Hall

C.2 Inclusion and exclusion criteria

The criteria for inclusion and exclusion

Inclusion	
Normal, healthy individuals over the age of 18, regardless of race, gender, religion, income, language or education.	
Vision must be normal, or corrected to normal (Libet *et al.*, 1983). Participants will be excluded if they require spectacles during testing as these can introduce noise into the eye tracking data, due to uncontrolled reflections off the glass (iMotions, 2018).	
Medications that do not alter EEG activity	
Atypical anti-psychotic medications	Risperidone, used in the treatment of bipolar disorder (Schomer & Lopes da Silva, 2011).
Antidepressant medications	New generation antidepressants have not been found to alter EEG activity (Schomer & Lopes da Silva, 2011). Examples include selective serotonin reuptake inhibitors (e.g. Prozac, Luvox CR, Pexeva (Fookes, 2018a)) and serotonin-norepinephrine reuptake inhibitors (e.g. Effexor XR, Pristiq (Fookes, 2018b)).
Analeptic (CNS* stimulant) medication	Methylphenidate (Ritalin/ Concerta), most commonly used in attention-deficit hyperactive disorder (ADHD) mostly has no change on the EEG (Banoczi, 2005). (*central nervous system)
Antibiotic medication	Amoxicillin trihydrate used for the treatment of some gram-positive and gram-negative cocci and rods have only been found to induce seizures in the event of an overdose. Only in the unlikely case of a suspected overdose, should the participant be excluded (and medical attention-sought immediately). (Banoczi, 2005)
Exclusion (these populations are excluded on account of various changes possible in the EEG)	
Migraines (at time of study)	Active migraines will present on an EEG, altering the signals. However, an untreated headache is not grounds for exclusion, as this will not be evident on the EEG. (Schomer & Lopes da Silva, 2011)
Previous head injurie s	Concussions (traumatic brain injuries) sustained within a 12-month period prior to the study will be grounds for exclusion. Literature comparing normal populations to those with concussions, with participants being excluded from the normal, or control groups if they had sustained a concussion within either a three month (Broglio *et al.*, 2011), six months (Broglio *et al.*, 2011; Yang *et al*., 2017) or 12 month period (Powers *et al*., 2014; Teel *et al.*, 2014) prior to the study, to avoid the long lasting effects on EEG activity.

Neurological conditions	History of epilepsy and seizures (Schomer & Lopes da Silva, 2011): epileptic seizures during infancy can result in altered EEG after the age of 2 in the non-ictal phases. Epilepsy is a contraindication in using the Tobii Eye Tracker 4C that will be used in this experiment, according to the Tobii safety guidelines. This altered EEG is generally a slowing of background activity and higher voltages as compared to normal patients. This can be due to the history of seizures, but can also be a result of the underlying pathological factors causing the epilepsy as well as the anti-epileptic medication themselves [typically phenytoin and carbamazepine (Salinsky *et al.*, 1995)]. (Konishi *et al.*, 1995) Cerebral palsy: altered brain EEG signals (Daly *et al.*, 2014) [aetologies including, but not limited to: asphyxiation, genetic/ metabolic disorders, infections etc. (Schomer & Lopes da Silva, 2011)]. Tumours have been found to alter EEG > 96 % of the time. Typical presentation are a slowing of activity and abnormalities such as dysrhythmias. (Schomer & Lopes da Silva, 2011)
Inflammatory CNS conditions	E.g. active encephalitis and meningitis, regardless of the aetiology (including, but, not limited to bacterial or viral infections, including HIV-related) have been shown to alter EEG patterns, mainly slowing of the EEG activity (Schomer & Lopes da Silva, 2011).
Caffeine intake on the day of the trial	Caffeine (1,3,7-trimethylxanthine), has been found to attenuate the power of α and β waves (Barry *et al.*, 2008; Meng *et al.*, 2017; Sieppman & Kirch, 2002). Caffeine is found in coffee, tea and caffeinated soft drinks as well as some medications such as those for asthma, hypersomnia as well as analgesic combinations (it is commonly found in migraine medication) (“Caffeine”, n.d.).
The following classes of medications have been found to alter EEG signals. Outliers are noted in the inclusion section	
Epinephrine	Epinephrine is an emergency medication (Multum, 2018) used in the event of anaphylactic shock and/ or cardiac arrest. A side effect of epinephrine is pupil dilation and will therefore affect the eye tracking. (Tobii Pro, n.d.)
Antipsychotic medication	Atypical antipsychotic: Clozapine [used in the treatment for schizophrenia(“Clozapine”, 2018; Schomer & Lopes da Silva, 2011). Typical antipsychotic: Haloperidol, is used in the treatment of schizophrenia and Tourette’s syndrome (speech and motor tics) (Amann *et al.*, 2003).
Antidepressant medications	Tricyclic antidepressants cause instability in frequency and voltage, with slowing of the α frequency (Schomer & Lopes da Silva, 2011).

	These are also still used as analgesics and in the treatment of sleeplessness. Examples include: imipramine (Tofranil-PM), amitryptiline (Elavil), doxepin (Sinequan), desipramine (Norpramin), nortriptyline(Pamelor) and protriptyline (Vivactil). (Fookes, 2018c)
Anxiolytic medications	Benzodiazepines have been shown to decrease α activity, and increase activity in the theta frequency range [4 – 8 Hz] (Schomer & Lopes da Silva, 2011). Examples include, but aren't limited to: alprazolam (Xanax), diazepam (Valium [high amplitudes of all frequencies (Banoczi, 2005)]), oxazepam (Serax), flurazepam (Dalmane), lorazepam (Lorazepam Intensol), trialozam (Hacion). (Fookes, 2018d)
Antihistamine medication	Diphenhydramine HCl** (Benadryl) has been shown to a general slowing of neural activity, mainly due to its sedative effects, which is represented on the EEG (Banoczi, 2005). **hydrochloride
Analgesic medication	Meperidine (Demarol) is a narcotic (opiate-based medication) that is used for the treatment of moderate to severe pain and causes an increase in the amplitudes of the EEG, as well as a slowing of α frequency signals (Banoczi, 2005). Opiate- based medications have been found to alter eye tracking data (Tobii Pro, n.d.).
Anticoagulant medication (therapeutic doses)	Warfarin sodium (Coumadin), used for prophylaxis / treatment of deep venous thrombosis, particularly after long periods of immobilisation of a limb (e.g. leg in a cast) has been found to be linked to the slowing of α frequencies as well as background neural activity (Banoczi, 2005).
Antihypertensive medication	Methyldopa (Aldomet) is an anti-hypertensive (control against high blood pressure) medication that has been found to cause slowing of alpha waves and general slowing of background neural activity (Banoczi, 2005). Anti-hypertensive medication has an effect on eye tracking (Tobii Pro, n.d.).
Chemotherapy and radiation	Active oncology treatment has been found to cause slowing of EEG patterns in some cases (Schomer & Lopes da Silva, 2011).
Recreational use of the following drugs will be grounds for exclusion on account of varying changes in the EEG	
Occasional use of marijuana (most active compound being THC: delta(δ)-9-tetrahydrocannabinol) will not influence the EEG, however, chronic use (including medical marijuana) has been shown to alter EEG signals (decrease in the power of α and β power bands, even after one month of abstinence (Schomer & Lopes da Silva, 2011).	
LSD (hallucinogenic [serotenergic psychedelic]) - lysenigic acid diethylamide (Banoczi, 2005)	

Cocaine (CNS stimulant) - methyl (1S,3S,4R,5R)-3-benzoyloxy-8-methyl-8-azabicyclo[3.2.1]octane-4-carboxylate (Heming *et al.*, 1994).
MDMA[stimulant and psychedelic] ("ecstacy"; 3,4-methylenedioxymethamphetamine) taken on its own, as well as in conjunction with THC will induce changes in the EEG (the latter causes different changes to MDMA taken in isolation] (Lansbergen *et al.*, 2011). Alcohol taken within two hours of the experiment will induce changes in the EEG (Ehlers, *et al.*, 1989) , however, it can be noted that simultaneous intake of MDMA will negate these changes (Lansbergen *et al.*, 2011).

Appendix D EEG data analysis

D.1 Checklist for EEG pre-processing

The following checklist was completed for each participants data during the pre-processing.

EEGLAB		
Action	Check	Notes
Import file: Using EEGLAB functions and plugins: From Brain Vis. Rec. .vhdr file		
Read in channel locations		
Filter the data 1.5 Hz and 50 Hz		
Run Cleanline to remove line noise (50Hz for SA)		
Append/ insert Cz Look up location – select as reference for the data Reference index 1:128		
Re-reference to average reference – select "re-add reference into the data"		
Change sampling rate to 256Hz.		
Visual inspection of channels: - Change amplitude to 50µV - Settings → time range to display (20 seconds) - Go through data and reject channels not showing EEG data > 60-70 % of the time		
Re-reference (average)		
Remove timestamps: 4 seconds before S64 and 5 seconds after S101/S102		
Remove artefacts from the data		
Eegh; - Timestamps of where data is rejected.		

- Copy paste into function line of excel Participant 1; rejection 1 (paste first timestamps under here); rejection 2 (paste second timestamps under here)		
Run AMICA		
Run DIPFIT Tools → plot head model BESA Manual co-registration → warp montage. Press okay Tools → Autofit. Save as participant_dipfitdone		
Plot → channel activations (scroll) → remove sections that are artefactual and may not require an entire channel for the decomposition Tools → reject components by map Excel table headings: components; scalp map dipole., artefact oscillations, frequency activation, time course pattern, DIPFIT		
Run AMICA again for comparison		
Run DIPFIT Tools → Autofit. Save as participant_2nddipfitdone Tools → DIPFIT → plot Select: Plot closest MRI slide, Plot dipole's 2D projections and plot projection lines		
Plot → channel activations (scroll) Tools → reject components by map Excel table headings: components; scalp map dipole., artefact oscillations, frequency activation, time course pattern, DIPFIT		

Remove components		
Extract epochs:		
General notes regarding the EEG pre-processing:		
Preparation for machine learning		
SET to CSV – run Python script		

Input to CNN model	Epochs	Results

D.2 Re-referencing techniques used in EEG analysis

The online referencing of the electrodes is to Cz. This is the most cephalic point of the electrode set up and a common reference point (Chella *et al.*, 2016) which is an advantageous position as it is surrounded by all the active electrodes (Teplan, 2002). However, in the case of this research we are interested in the information around Cz, as Cz measures neural activity related to executive motor function. Thus, it is important that the information at the online reference. Other physical references include ipsilateral – or contralateral ears as well as linked ears or linked mastoids (Teplan, 2002). The digitally linked mastoid technique is an offline referencing technique (offline means the reference is changed post data collection; while online refers to referencing during data collection). In the digitally linked mastoid technique there is the creation of a "linked" reference which is the average between the two electrodes placed at the mastoids (TP9 and TP10). However, these electrodes tend to be noisy as it is difficult to

achieve- and then maintain throughout the experiment - adequate impendence levels. This is due to the cap's sometimes poor conformity to different head shapes – making contact difficult at certain electrode sites. Further, hair tends to be thicker in this region thus contributing to poor electrode contact and increased impedance (resistance). The physical reference also relies on the assumption that the impedance levels in the reference electrodes remain constant throughout the experiment. There are also non-physical reference points (i.e. "reference free") which can mitigate the cons describe above. "Reference free" techniques discussed next.

"Reference free" techniques are references without a physical electrode. These include the average reference (also referred to as the common average reference) where the EEG potentials are all referenced to an average of all the EEG potentials (Chella *et al.*, 2016; Teplan, 2002). This choice is appropriate to this research as there is an even distribution of electrodes across the entire scalp (as opposed to increased density of electrodes on a single lobe) Another technique is the Reference Electrode Standardisation Technique (REST) in which a virtual reference is located at infinity (Chella *et al.*, 2016). There is no overt disadvantage for choosing a reference free technique over a physical reference, however Dien (1998) suggests the average reference may be flawed in its technique in that it requires the assumption that the bottom half of the head is sampled the same as the top half – it is important that these techniques are only used when there is an even distribution of electrodes (and not a concentration over a certain brain area)..

It is important to use a reference that can be used for comparison with other studies (Dien, 1998).

D.3 The theory behind independent component analysis

In contrast principle component analysis (PCA) is aimed at dimensionality reduction and minimisation of the data to a single component whereas ICA aims for the identification of all source components. The mathematics behind ICA will be described below. Certain pre-requisites need to be met and assumptions need to be made in order for the ICA decomposition to be considered reliable. These are discussed next.

The following pre-requisites need to be met:

1. The data must be continuous (Delorme & Makeig, 2004)
2. There must be an adequate number of data points (D) for the temporal independence of the underlying sources of components to be decomposed (Delorme & Makeig, 2004). In order to meet this pre-requisite, all data must be collected as a continuous EEG data matrix across the whole experiment. There is discrepancy in the literature as to what is "enough" data. Two different sources recommend (Delorme & Makeig, 2018; Gabard-Durnam *et al.*, 2018) a different number of minimum data points. The standard formula is presented as follows:

$$D = N\ x\ (number\ of\ components)^2$$

Where number of components = number of channels

Equation 1

The discrepancy lies in the value for N:

$$\boldsymbol{D = N\ x\ (128)^2\ at\ 256Hz}$$

If $N = 3$ (Delorme & Makeig, 2018)

$$D = 491\ 52\ \frac{samples}{256\ Hz}$$

$$D = 192\ s = 3.2\ minutes$$

If $N = 30$ (Gabard-Durnam *et al.*, 2018)

Equation 2

In terms of this research - there is little benefit in excluding a participants data should their continuous data matrix fall short of 32 minutes (Gabard-Durnam *et al.*, 2018). This research is exploratory, and the focus is not in source space analysis – where each independent component of cognitive data is analysed in isolation and used to make inferences about brain functioning. The ICA is used as a processing step for data cleaning in preparation for machine learning and not for the individual cognitive component analysis. All datasets used contain between 15 and 40 minutes of continuous data.

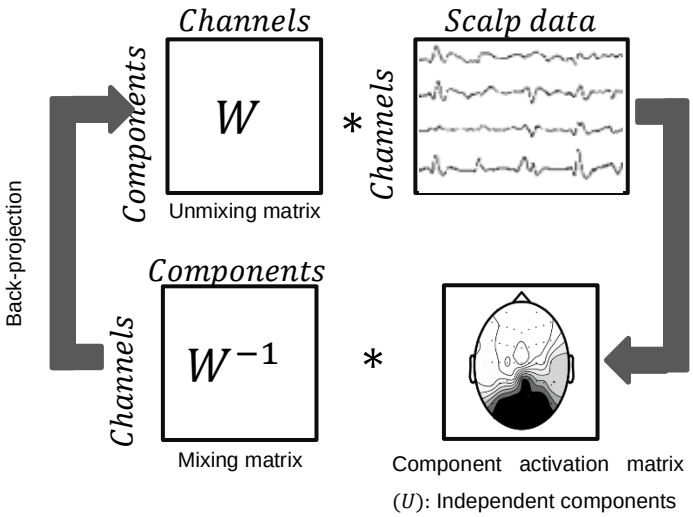
Further assumptions are made regarding the ICA decomposition:

3. Sources are localised (i.e. won't find eye blink components in the occipital area) and independent (the source of the eye blink is not the same as a muscle twitch nor as a neural process); assuming the complexity of the brain can be modelled as such (Ahirwal & Londhe, 2012; Makeig *et al.*, 1996)
4. Volume conduction in the brain is effectively instantaneous – i.e. propagation delays between the arising neural electrical potential and the electrode are negligible (Makeig *et al.*, 1996)
5. Assumption that $N_{sources} = N_{components}$, where N refers to the number thereof. This one is limited in the basis of its assumption in that $N_{sources}$ can exceed $N_{components}$ to any degree
6. Sources of neural components are not Gaussian, and something such as line noise can be considered sub-Gaussian (Delorme & Makeig, 2004)

These assumptions do have their limitations, but the limitations are such that the ICA decomposition is still reliable, when these are taken into account simultaneously.

The basis of ICA is to find a weight matrix (W) such that the source components (X) are statistically independent. ICA forms a series of spatial filters which, when passed over the data, cancel out all but one of the relevant distinct source signals that constitutes a part of the multichannel data. These filters are learned (by the ICA decomposition algorithm) and each has a focus for a particular source signal. Formally, this is a linear decomposition. A signal matrix is composed of any weighted sum of the component signal matrices.

The following diagram will be used to describe this process in detail:



A flow diagram outlining the process of ICA. Adapted from the following sources: ("How to Read an EEG", n.d.; Makeig & Onton, 2011; Zhang et al., 2004)

The unmixing matrix (W) of size $components\ x\ channels$ is multiplied by the data matrix (size $channels\ x\ scalp\ data$). The unmixing matrix is the set of spatial filters learned by this process of ICA decomposition. This gives us the independent components, defined as the portion of the entire multichannel recorded dataset when separated from the remaining recorded data (Makeig & Onton, 2011). The independent component maps give us the activation time courses of these components and the scalp map above is an example of what the EEGLAB interactive window will show in colour. These scalp maps remain constant over time as the EEG is considered spatially stable (Makeig & Onton, 2011). The time series (not depicted) will give the amplitudes and polarity (+/-) at each time point.

The component mixing matrix (W^{-1}) is multiplied with the independent components to reconstitute the data from the scalp map, through the process of back-projection.

Algebraically this is represented as follows:

$$U = WX$$

$$X = W^{-1}U$$

Equations 3;4

The notations used in the above formulae are defined as follows:

Notation used in the ICA decomposition (Equations 3;4)

Notation	Explanation
U	Component activation matrix (two parts: component scalp maps as well as time series)
W	Unmixing matrix of spatial filters learned by ICA decomposition
X	Scalp data matrix
W^{-1}	Component mixing matrix: Columns give the relative strength and polarities of the projections of one component source signal to each of the scalp channels.

(Makeig & Onton, 2011)

D.4 The features used to analyse independent components in EEG data

Feature	Explanation
Scalp map dipole	This gives an indication of the location of the strength of the dipole (an unclear dipole would suggest a non-cognitive source The strength and clarity of the dipole is indicated by the intensity and range (blue to red) of colours. The stronger and clearer the dipole, the closer to the surface the dipole is, and the more intense the colours will appear ICA cannot reliably decompose sources in the deep brain (i.e. the thalamus). The closer the source of the dipole is to the surface, the stronger and clearer the dipole. A lesser range and intensity of colours can be indicative of an electromyographic (EMG) source, which is not dipolar
Artefact oscillations	This is a green field distribution which gives an indication of where the source arose in the time-series of the experiment. A more evenly distributed presentation suggests a cognitive component, as unlike EMG, cognition is not intermittent (however, a particular cognitive component could be more prominent within the distribution according to the demands of the task) Stronger cognitive processes appear denser, with an even distribution of red and blue spots. These typically arise closer to the surface. Deeper sources will have an even distribution, but appear sparser. EMG will appear in bands – the density dependent on the strength of the source
Frequency activation	This is otherwise known as the activity power spectrum. It is represented in a graph format which gives an indication of the frequency (or different frequencies) within a single channel. Cognitive sources will have prominent

	peaks corresponding to their respective frequency ranges. For example, an α source will peak at 8 – 10 Hz, or 10 – 12 Hz. EMG will have a consistent rise in the power spectrum graph, whereas line noise will have a distinct flat line
Time -series course	This the actual component scroll data which gives a visual representation of the components change in frequency through the time course of the experiment Cognitive data has distinct wave forms which can be recognised, and peaks can be counted in one second windows. Cognitive data with irregular and recurring artefacts should be considered for removal. This is a trade-off, as cognitive data will be lost, but the artefacts could skew the rest of the data if left in the matrix EMG has a consistent high frequency with extremely variable amplitude presentation. ECG can also be found as a source component. This appears as regularly spaced sharp increases in positive amplitudes. Artefacts tend to have sharp changes from positive to negative peaks, and appear irregularly. They can be recurring or non-recurring Eye blinks and lateral eye movements have stereotypical patterns that can be easily identified for rejection
DIPFIT (dipole location)	This uses a built-in function of EEGLAB to present (with a calculated percentage of confidence – the residual variance [RV]) the component source in its estimated position on an MRI scaled brain, which takes the channel locations read in in as well as the source decomposition into account. This is a useful tool that can help clarify whether a component is artefact or cognitive based on its source location

Appendix E Deep learning analysis

E.1 MinMax Scaler

The formula for the MinMax scaler used to prepare the data for the CNN model.

$$n = \frac{(x - x_{min})}{(x_{max} - x_{min})}$$

Equation 1

The following table describes the notation used above:

The notation used in the MinMax scaler equation

Notation	Explanation
n	New data point
x	Original value
x_{min}	Minimum value in the original data
x_{max}	Maximum value in the original data