Author/s:
Patel, Giana Rose

Title:
Using long-wear electroencephalography to ascertain the variability of Lempel-Ziv Complexity (LZc) measures of consciousness

Date:
2019

Persistent Link:
https://hdl.handle.net/11343/240608

Using long-wear electroencephalography to ascertain the variability of Lempel-Ziv Complexity (LZc) measures of consciousness

Giana Patel

ORCID: 0000-0003-4627-3074

Thesis submitted in total fulfilment of the requirements of Master of Philosophy (Psychological Sciences)

December 2019

Faculty of Medicine, Dentistry, and Health Sciences

Melbourne School of Psychological Sciences

The University of Melbourne, Australia

Abstract

It has been recently claimed that measures of spontaneous electroencephalography (EEG) signal complexity, such as Lempel-Ziv Complexity (LZc), can provide an index of an individual's level of consciousness. Research and clinical practice are currently limited to unreliable behavioural and physiological measures to indicate consciousness. Therefore, there is significant urgency for an objective, reliable, brain-based measure of consciousness. EEG complexity measures utilise algorithms from Information Theory to quantify the diversity in spontaneous EEG data. These are being used to measure the diverse neural activity which necessarily underlies conscious experience. LZc assesses the complexity of multi-channel EEG data using a compression algorithm. Studies of LZc typically involve comparing conditions of altered consciousness with periods of conscious wakefulness. These studies suggest that the change in complexity observed is reflective of the change in level of consciousness. However, very little is known about how LZc varies, either with or without a corresponding change in consciousness.

The present study utilised portable long-wear EEG to record multi-day, continuous EEG data from two participants (a total of 8 days for Participant 1 and 4 days for Participant 2). Data from each participant was analysed independently. A LZc algorithm was used to compute a complexity value for every non-overlapping 10-second segment. Results demonstrated that, as with previous research, LZc during Wake (14-hours during the day, multiple days per participant) is, on average, higher than during sleep (Stage N1, Stage N2, Slow-Wave-Sleep, and REM sleep). However, there is considerable variation surrounding these means. Visualising LZc across Wake revealed a consistent but wide spread of variability around the mean, with a scattering of low LZc values reflected by a negative skew in the data. This also results in a wide range of possible mean LZc values made available from taking samples (between 1 and 120 minutes in duration) during this period. Although this variability reduces with larger samples sizes, even day-to-day, LZc can significantly differ within a person.

Regardless of the source of this variability, its presence causes concern due to the overarching clinical motivations and potential practical applications of this measure. These results suggest that LZc may not be indicative of level of consciousness, as previously claimed. The issues raised and addressed in this study are not unique to LZc, but will apply to all complexity algorithms, current and future. With this study, we have shown that long-duration EEG is a successful framework for identifying variability in a complexity measure of consciousness. This information-rich dataset is uniquely capable of exposing and investigating complexity measures, with the additional insight of observing and analysing complexity across time. This study endeavours to redirect discussions of this field and promote the use of this framework to both acknowledge and empirically address all surrounding issues and assumptions. All complexity measures should undergo reliability testing as both a proof of concept and a proof of practice before being utilised in research or clinical applications.

Declaration

This is to certify that:

i. The thesis comprises only my original work towards the master of philosophy except where indicated in the preface,

ii. Due acknowledgement has been made in the text to all other material used,

iii. The thesis is fewer than 50,000 words in length, exclusive of tables, maps, bibliographies and appendices, as approved by the Research Higher Degrees Committee.

Giana Patel

709533

20/12/2019

Preface

I acknowledge Seer Medical and St Vincent's Hospital, Melbourne for their involvement in data collection, providing us with the long-duration EEG data which was analysed in this study.

Acknowledgements

Thank you to my supervisors, advisory committee, and extended staff and students at the University of Melbourne who have been here with me throughout my degree. Specifically, I would like to deeply thank my supervisor, Associate Professor Olivia Carter. You have provided me with both expertise as well as guidance, and I would not have been able to finish this thesis without your support and encouragement. To my co-supervisors, Dr Christian Nicholas and Dr Hinze Hogendoorn, thank you for your continued interest and dedication to this project. You have each given a lot of time and expertise; I have learnt a lot from you both.

To my friends and family, thank you for your patience. The support I have been given over the past two years, especially these past few months, has allowed me to completely immerse myself in this thesis and produce something I am proud of. I would like to thank my parents, Tracie and Raman, and my brother, Levi. To Fitzy, thank you for encouraging me to spend time outside in the sunshine; you have kept me sane. Lastly, to my friend Nicole, you are an endless inspiration to me and my academic pursuits, and I strive to do my best work because I am inspired by you and your dedication. I would not be where I am today without the people around me.

Table of Contents

List of Figures

List of Tables

Introduction

Consciousness is something we are all intimately familiar with from our own subjective experience of the world, yet we can only currently infer consciousness in other people via the recognition of purposeful action and behaviour (Chalmers, 1995). However, a person does not need to interact with their environment in this way for conscious experience (Koch, Massimini, Boly & Tononi, 2016; Sanders, Tononi, Laureys & Sleigh, 2012). Many instances exist where one may be disconnected or unresponsive to their external environment yet retain some degree of internal conscious awareness. Our ability to identify altered or reduced signs of conscious awareness has significant implications, particularly for patients in clinical care. For example, an absence of signs of consciousness, alongside severe neural injury, may be used as evidence for the withdrawal of life-supporting medical care or impact decisions regarding ongoing care and pain management for these patients (Seel et al., 2010). Currently, the clinical assessment of consciousness is predominantly limited to behavioural assessments (Owen, 2008). Therefore, there is a significant need for an objective and reliable measure that can indicate an individual's level of consciousness.

Current theoretical models posit consciousness to be complex, meaning neural activity must be differentiated and integrated (Tononi, 2004). Algorithms that assess for the diversity of electroencephalography (EEG) signals have been proposed as measures of this diverse neural activity which underlies conscious experience. These are referred to as dynamic complexity measures of consciousness. One example of such measures is Lempel-Ziv complexity (LZc), which utilises a Lempel-Ziv compression algorithm (Lempel & Ziv, 1976) to analyse the EEG data. It has been observed to successfully index levels of consciousness in different groups (Schartner et al., 2015; Schartner et al., 2017a; Schartner, Carhart-Harris, Barrett, Seth & Muthukumaraswamy, 2017b). However, it is unknown how this measure may vary, either with or without a corresponding change in consciousness. Given the significant clinical implications of such a measure, a thorough investigation of its reliability is imperative. The present investigation aims to describe the variability of LZc, in order to test the reliability of this tool. This will serve to further establish a framework for testing other complexity measures before they are implemented as a measure of consciousness.

Neuroscience of Consciousness

Consciousness is commonly conceptualised as our subjective experience of the world as individuals. It is a cumulative awareness of the self and our environment, supported by patterns of neural activity and arousal in the brain (Pereira & Ricke., 2009). This cumulative conscious experience is often thought to vary along a continuum from complete unconsciousness to vivid wakefulness (Seth, Dienes, Cleeremans, Overgaard & Pessoa, 2008). However, there is still some debate regarding the nature of this spectrum (Overgaard, Rote, Mouridsen & Ramsøy, 2006; Overgaard and Overgaard, 2010; Fingelkurts, Fingelkurts, Bagnato, Boccagni & Galardi, 2014). Studies considering subjective reports of participants' conscious perception of a stimulus often present consciousness as an all-or-nothing phenomenon (Koch et al., 2016). However, neuroscientific studies exploring the neural activity associated with these perceptions suggest that conscious awareness is more resemblant of a continuum, with different levels of awareness relating to different levels of representation in the brain (Grill-Spector, Kushnir, Hendler & Malach, 2000; Bar et al., 2001; Kouider, de Gardelle, Sackur & Dupoux, 2010). Much of the neural activity associated with conscious experience may not pass the threshold to allow verbal report but still contributes to an individual's conscious experience (Dehaene, Changeux, Naccache, Sackur & Sergent, 2006; Kouider et al., 2010). Different global states of consciousness (e.g. awake or asleep) are commonly referred to as different levels of consciousness. These are periods of relative stability that vary along a scale (Tononi, 2008). However, the concept of levels implies an ordering of higher or lower consciousness (Bayne & Carter, 2018) and it has been argued that consciousness should be theoretically described across multiple dimensions instead (Bayne, Hohwy & Owen, 2016).

The neuroscience of consciousness has long worked towards identifying the neural mechanisms underlying the phenomenological experience of conscious awareness. This includes investigations into the content that reaches conscious awareness, and also the neural activity that supports consciousness itself (Koch et al., 2016). The neural mechanisms responsible for this overall consciousness combine to support a holistic, emergent experience, irrespective of specific content (Koch et al., 2016). Consciousness is the connection of cortical and subcortical networks interacting (Sarà et al., 2011) as opposed to specific activity localised to a region of interest. Sarà and colleagues (2011) describe this correspondence of networks as a complex system with many variables and a hierarchy of systems controlling the output. Identifying one aspect of the system does not inform the functioning of the whole (Sarà et al., 2011). For example, identifying the neural activity associated with a conscious perception does not provide insight into the individual's overall conscious experience. Therefore, to identify and investigate global states of consciousness, one must identify characteristics of complex but interconnected neural activity as they occur throughout the brain.

This field of research aims to determine and measure neural activity that may be associated with various global states of consciousness. Neuroimaging and neurophysiology techniques, such as EEG and functional magnetic resonance imaging (fMRI), are employed to identify and quantify the neural correlates of global states of consciousness (Koch et al., 2016). These neural correlates of consciousness (NCC) are typically pinpointed and investigated by contrasting global characteristics of neural activity during conscious wakefulness against that of an altered state of consciousness. This approach attempts to identify changes in neural activity associated with changes in consciousness, independent of changes in physiological arousal (Schartner et al., 2017b). This can be done by comparing results between participant groups with different levels or states of consciousness, manipulating the level of consciousness or inducing a loss-of-consciousness using drugs, or by contrasting wakefulness against the stages of sleep.

Clinical Assessments of Consciousness

In many circumstances, it is possible to be conscious yet disconnected from the environment, with an inability to report on your conscious awareness. Examples include dream-like experiences under ketamine anaesthesia (Sarasso et al., 2015), detached consciousness during absence seizures (Bayne, 2011), limited sensory and motor function in patients with Disorders of Consciousness (DOC) and locked-in syndrome (Bruno, Vanhaudenhuyse, Thibaut, Moonen & Laureys, 2011), or simply while dreaming during sleep (Siclari et al., 2017). Research and clinical practice are currently largely reliant upon the identification of behaviours and subjective reports to indicate conscious awareness (Giacino, Kalmar & Whyte, 2004). This is restrictive because it requires the person to interact with the environment in certain ways to provide evidence of their awareness. Therefore, the need to develop objective measures that reliably index consciousness in both clinical and experimental settings is a priority.

Anaesthesia. Anaesthesia is an example of an everyday, clinical manipulation of consciousness in which patients are sedated via general anaesthetic drugs when undergoing medical procedures (Kissin, 2000). An anesthesiologist will use a combination of physiological measures, knowledge of drug dosages, and behavioural observations to ensure the patient remains unconscious for the duration of the procedure (Kissin, 2000). Many EEG-based methods of monitoring sedation levels have been proposed. Typically, these analyse EEG signal from a single electrode on the patient's scalp during sedation and aim to provide an indication of the patient's overall level of consciousness using a single numerical measure (Voss & Sleigh., 2007). The bispectral index (BIS) is one of the only EEG-based methods implemented into common clinical practice (Kissin, 2000). EEG data is recorded and analysed by the BIS from electrodes on the patient's forehead and the level of sedation is reported in real-time. BIS scores are given on a scale from zero (flat-line EEG) to 100 (fully alert and awake) (Sleigh & Donovan, 1999). The BIS was empirically developed by using anaesthesiologist's classifications of sedation levels, alongside EEG recording data, with specific profiles developed for specific anaesthetic drugs (Sleigh & Donovan, 1999). However, this was purely empirical and not based in theory of consciousness or anaesthetic sedation. Additionally, information regarding how the BIS analyses the EEG data to compute this level of sedation has never been publicly released. Due to the lack of understanding about its operation, it is generally unsupported by anaesthesiologists and neuroscientists, and is not thought of as a reliable measure of consciousness (Ontario, 2004). Multiple alternative means of monitoring sedation levels using EEG have been proposed. However, none have proven to be more useful than the physiological and behavioural assessments that are currently used (Voss & Sleigh, 2007).

Disorders of Consciousness (DOC). The lack of an objective measure of consciousness is also a significant problem for the clinical care of patients with DOC. DOC result from severe neural injury and involve substantially reduced and altered levels of conscious awareness and wakefulness in an individual (Bruno et al., 2011; Owen, 2008). DOC diagnoses vary along a single dimension related to signs of wakefulness and awareness. Clinical assessment and diagnosis are based on the behavioural assessment of these signs (See Table 1). These signs include eye-opening, command following and intentional movements (Giancino et al., 2004; Owen, 2008; Seel et al., 2010). Diagnosing and differentiating between Coma, Unresponsive Wakefulness Syndrome (UWS – previously Vegetative State), and

Minimally Conscious State (MCS) (see Table 1.) is a significant challenge (Owen, 2008). Patients with DOC often exhibit reduced sensory and motor skills, alongside cognitive impairments (Schiff, Nauvel & Victor, 2014), which make behavioural assessments of their cognitive condition particularly challenging.

Table 1. *Description of the main classifications of DOC and accompanying behavioural characteristics used for diagnosis (Adapted from Schiff, Nauvel and Victor, 2014)*

Syndrome	*Definition*	*Behavioural Characteristics*
Coma	State of complete unresponsiveness, no signs of wakefulness or awareness	Eyes-closed, unresponsive to environment and pain, no intentional movement, often temporary
Unresponsive Wakefulness Syndrome (UWS)	State of intermittent arousal, signs of wakefulness but no awareness - Previously known as Vegetative State (VS)	Eyes-open, signs of a sleep-wake cycle, reflexive movements of eyes and limbs, no evidence of intentional or goal-directed behaviour
Minimally Conscious State (MCS)	State of wakefulness and fluctuating signs of awareness	Intermittent signs of responsiveness to verbal commands, intentional behaviours, purposeful eye movements, may verbally communicate. Patients vary in the degree and frequency of signs of awareness, fluctuating in and out
Locked-in Syndrome (LIS)	State of complete loss of motor function but with full awareness retained - Not a DOC	Complete loss of motor function, aside from eye-movements. Full awareness retained. Often misdiagnosed as a DOC due to no physical signs of awareness and the inability to communicate

Due to these challenges, rates of misdiagnosis of DOC patients are high. This is particularly problematic when distinguishing between UWS from MCS, where diagnostic error rates are estimated up to 45% (Schnakers et al., 2009). Additionally, recent evidence reveals that some UWS patients retain an ability to understand and respond to commands, despite being unable to show behavioural evidence of this. An fMRI study by Owen and colleagues (2006) provided spoken commands to participants and instructed them to use mental imagery to respond. Yes or no answers were given by the participants by either imagining walking through a house or imagining playing tennis. These yes or no responses were visible to researchers due to the distinctly different patterns of neural activity produced by the two mental imagery options. Results revealed that a portion of patients diagnosed as UWS were actually more consciously aware than behavioural assessments suggested (Owen et al., 2006). Similar studies have also used motor imagery instructions to communicate with patients using EEG (Cruse et al., 2011; Goldfine, Victor, Conte, Bardin & Schiff, 2011). Neuroimaging studies such as these are providing a unique insight into the internal world of DOC patients and highlighting the ethical significance of accurate assessment methods.

Current diagnosis methods are largely inadequate due to their reliance on doctors formulating subjective decisions based on subtle behaviours (Bender, Jox, Grill, Straube & Lule, 2015; Giancino, Fins, Laureys & Schiff, 2014). The behavioural signs are challenging to identify and the distinction between them hard to define, yet these diagnoses are used to inform important decisions regarding patient care. The difference in behaviour between UWS and MCS is subtle but the difference in diagnosis is critical. With a misdiagnosis rate of up to 45% (Schnackers et al., 2009), this indicates that many patients may be more aware than they seem. This statistic raises a significant ethical concern in addition to an inherent diagnostic problem. These concerns harbour the possibility of clinical implications for medical treatment plans, such as pain management and life support. There are also ethical implications for patient care and wellbeing, including personal needs for social attention or entertainment. In states of reduced or altered consciousness, the ability to physically or verbally express awareness is often modified or removed (Koch et al., 2016; Sanders et al., 2012). Therefore, an empirical measure that can identify reduced levels of consciousness, without relying on sensory input or motor output is necessary. However, it is clinically and ethically critical that these are thoroughly tested and proven to be reliable before being implemented in practice.

Complexity of Consciousness

Complexity models of consciousness link together a theoretical explanation of consciousness with a mechanistic explanation for how this may occur in the brain. Complexity models propose the idea of conscious experiences being differentiated (composed of many parts and each experience distinct from the next) but also integrated (consciousness is experienced as a coherent whole) (Tononi, 2008). These models suggest that consciousness requires these theoretical properties to be reflected in the properties of neural activity (Seth, Izhikevich, Reeke & Edelman, 2006; Tononi, 2008). Therefore, neural activity is thought to be similarly differentiated (dynamically distinct subsets of information-rich neural activity) and integrated (distributed activity across the brain working as a whole, coherent system) in order to support conscious experience (Tononi, 2008). This neural diversity must, however, function within some optimal middle range, to balance a sufficient level of diversity with enough order to maintain efficiency (Carhart-Harris, 2018).

There is currently no consensus on a single model of consciousness. However, the complexity model of consciousness maintains growing support due to its association with many empirical attempts to quantify global states of consciousness. This model proposes that a certain level of differentiated neural signal is necessary to support a rich and diverse conscious experience, and more importantly, that this can be measured (Sitt et al., 2014; Tononi, 2008). Additionally, the model supports the proposal of long-range communication and coordination throughout the brain (Dehaene & Changeux, 2011; Lamme & Roelfsema., 2000; Tononi, 2008). A higher degree of neural complexity is expected when consciousness is present, compared to a lower level of complexity whenever consciousness is lost. Examples include states of sleep, anaesthesia or coma (Tonoi, 2004; Tononi, Sporns & Edelman, 1994).

Several empirical measures of consciousness have sought to quantify integration and differentiation in the brain. Casali and colleagues (2013) developed a method to assess both differentiation and integration with a single metric: the Perturbational Complexity Index (PCI). The PCI method uses Transcranial Magnetic Stimulation (TMS) to perturb the cortex and induce electrical activity which spreads throughout the brain (Massimini et al., 2007). This activity is subsequently recorded via EEG across the scalp, then a normalised Lempel-Ziv algorithm (Lempel & Ziv, 1976) is used to assess the differentiation of the signal received. The complexity of this induced signal, as measured by the Lempel-Ziv compression algorithm, constitutes the PCI score. In a healthy, waking participant, the TMS triggers a pattern of activity that is widespread and complex, and therefore, produces a high PCI value. This indicates a combination of high integration and differentiation, proposed to be supportive of conscious experience (Casali et al., 2013). A high PCI value will only exist if the signal is both integrated and differentiated (See Figure 1) (Casali et al., 2013; Sarasso et al. 2015).

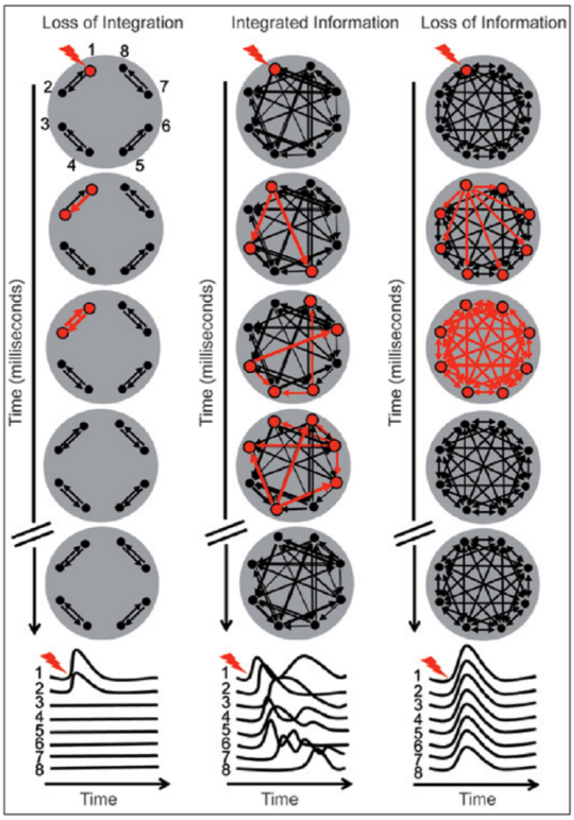

Figure 1. Integration and differentiation of neural activity, as measured by PCI. TMS is used to perturb the cortex (top) and the subsequent neural activity as measured by EEG (bottom). A lack of integration

(left) results in bursts of local activity with no widespread connection across the brain. A lack of differentiation (right) results in widespread activity, but this activity is not diverse. A combination of integration and differentiation (centre) results in widespread and diverse activity across time. This combination of integration and differentiation is required for a high PCI result (From Sarasso et al., 2014).

The PCI metric is interpreted as an objective scale, and researchers claim an ability to distinguish between different levels of consciousness within range of 94-100% sensitivity and specificity (Casali et al., 2013; Casarotto et al., 2016; Gosseries, Thibaut, Boly, Rosanova, Massimini & Laureys, 2014; Sarasso et al., 2014). However, this level of discrimination implies the presence of clear boundaries between different global states of consciousness and that a single measurement can differentiate between states. It also implies that it can be utilised to devise claims about a person's state. This simplification concerningly negates the importance of understanding the nature and fluctuations of reduced conscious experiences**.** The use of TMS also bears limited practicality. Since PCI requires a perturbation of the cortex, there lies an unknown level of risk of inducing seizures in brain-damaged DOC patients (Wassermann, 2000). This equipment is also expensive, impractical and uncommon in hospitals. Additionally, it would require other means of neuroimaging to first identify healthy areas of the brain to target the TMS (Casali et al., 2013; Gosseries et al., 2014). Therefore, there may be limited practical utility for PCI as an everyday clinical diagnostic tool.

While the development of measures for quantifying global states of consciousness crosses all forms of neuroimaging and neurophysiology, the field has gained recent momentum with EEG-based measures. EEG is a method of recording the electrical activity associated with neural activity, via electrodes positioned on the scalp (See Figure 2) (Tatum, 2014; Niedermeyer, Schomer & Lopes da Silva, 2010). Spontaneous electrical activity is recorded over time and can be analysed in multiple ways to learn about neural activity and cognition

(Niedermeyer et al., 2010; Tatum, 2014). EEG is non-invasive, low-cost, and mobile, allowing for bedside implementation. It is also already routinely collected in relevant medical cases, including DOC patients (Sitt et al., 2014). Due to its particularly fine temporal resolution and ability to record over long durations, it is uniquely suitable for applications of measuring consciousness. EEG also has the potential to provide insight into the functional nature of the brain across time (Neidermeyer et al., 2010). For DOC, this is highly beneficial, since signs of consciousness are often inconsistent and fluctuating (Bender et al., 2015). Moreover, EEG data can be analysed as it is being recorded. Therefore, it harnesses the potential to track changes and provide feedback on the patient's level of consciousness in real-time.

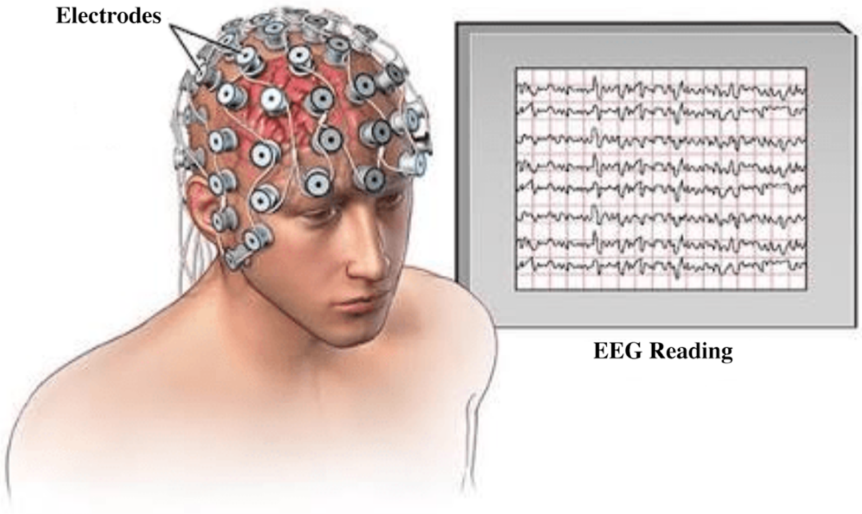

Figure 2. Image showing a typical EEG recording set up. Electrodes are placed onto the surface of the scalp, between parted hair (left). A conductive liquid, gel, or glue is used to hold them in place and maintain a connection. Each electrode is attached to a wire, connected via an amplifier system to a recording system on the computer. The monitor (right) shows an output of raw EEG recording, where each line represents the signal recorded from one electrode channel across-time (From Siuly, Li & Zhang, 2016).

The ability to record and analyse EEG without patient participation renders it ideal for this application of measuring consciousness. Many previous approaches to measuring consciousness have been limited in practicality and efficacy due to their reliance on behavioural or mental tasks (van den Brink, Nieuwenhuis, van Boxtel, Luijtelaar, Eilander & Wijnen, 2018). Any activity requiring higher-order cognition or physical participation is unsuitable for assessing patients where these functions may not be preserved (Brink et al., 2018). In any case of reduced or altered consciousness, interaction with the external environment is likewise reduced or altered (Sanders et al., 2012). Thus, a measure that can operate independently from these abilities will be far more useful and informative in this context of research. Over recent years, task-free paradigms have increased in popularity (Brink et al., 2018; Sitt et al., 2014).

EEG-Based Complexity Measures. Inspired by complexity models of consciousness and the development of PCI, recent attempts to develop an empirical measure of consciousness aim to measure complexity through characteristics of differentiation in spontaneous EEG signal. This differentiation is measured using algorithms from Information Theory. Information Theory refers to a field of mathematics and computer science that designs and implements algorithms to analyse, quantify, store and transmit information (Ghahramani, 2006). When applied to continuous EEG, data is fed through an analysis that utilises one of these algorithms to quantify the diversity of information contained in the signal. This is achieved by quantifying characteristics such as randomness, entropy, irregularity, or compressibility. A single number is generated to reflect the overall complexity of each segment of continuous EEG data. Unlike PCI, these do not assess the combination of integration and differentiation across a network. Instead, they solely measure signal diversity over time (Shartner et al., 2015). This approach allows for investigations into changes of whole-brain-level dynamics across time, and additionally, does not require the brain to be stimulated using TMS.

Research shows an association between low informational complexity and low consciousness. This is irrespective of the type of dynamic complexity algorithm used and regardless of whether consciousness is altered via pharmacological (anaesthesia), physiological (sleep), or brain trauma (DOC) mechanisms. Recordings produced during reduced consciousness are generally observed to be more regular than during conscious wakefulness. Hence, metrics of informational complexity are lower when the conscious level is lower (Blume, del Giudice, Wislowska, Lechinger & Schabus, 2015). An example of this occurrence is when the inhibitory neurotransmitter effects of anaesthetic drugs decrease the speed of neural transmission and increase the regularity of EEG signal, which is reflected in low measures of informational complexity (Wang et al., 2017). A comprehensive review by Sitt and colleagues (2014) explored 92 methods of analysing EEG data, to determine their usefulness in discriminating levels of consciousness in DOC patients. From these 92 methods, it was concluded that entropy and complexity-based algorithms were one of three general approaches that were most indicative of individuals' global states of consciousness. For example, the measure Kolmogorov-Chaitin complexity (K), a compression algorithm, was noted to be significantly higher in MCS patients compared to UWS patients. Both the average complexity and inter-trial stability increased as consciousness increased, suggesting that complexity can index consciousness in some clinically useful way (Sitt et al., 2014). While inspired by a complexity model of consciousness, these informational complexity measures are still empirically developed and only correlational in their relationship to consciousness. Although this evidence demonstrates promise to successfully index consciousness, it requires further in-depth investigation.

Lempel-Ziv Complexity (LZc). LZc employs a compressibility algorithm (Lempel & Ziv, 1976) to assess the informational complexity of EEG data (Figure 3). It does so by counting the number of unique components in a string of data, quantifying the number of distinct patterns of activity in the data segment (Schartner et al. 2017a). Originally developed by Lempel and Ziv (1976), it is a family of algorithms commonly used in computer science with many applications. For example, LZc is used for lossless data compression of digital information to create a zip file on a computer (Jain & Lakhtaria, 2016). Psychological research has applied this for many years to both EEG and fMRI data, particularly in early depth-of-anaesthesia monitoring research. LZc was first utilised with spontaneous EEG to measure levels of consciousness in 2001, by Zhang and Colleagues. LZc levels were observed to reduce during anaesthetic sedation compared to conscious wakefulness. It performed more effectively than other algorithms at discriminating between conscious and unconscious anaesthetic states, performing with up to 93% accuracy (Zhang et al., 2001).

The LZc algorithm is a simple computation of the compressibility of information contained in an EEG signal. It holds an extensive history of applications in psychological research and is experiencing a stark rise in popularity. For this reason, we have chosen to focus this discussion on LZc, as an example of a complexity measure of consciousness. However, given that these complex algorithms are all similar in concept, application, and findings, the critiques brought against LZc will be of concern to other algorithms as well. Additionally, discussions of ways to address concerns surrounding LZc will apply similarly to other EEG complexity algorithms.

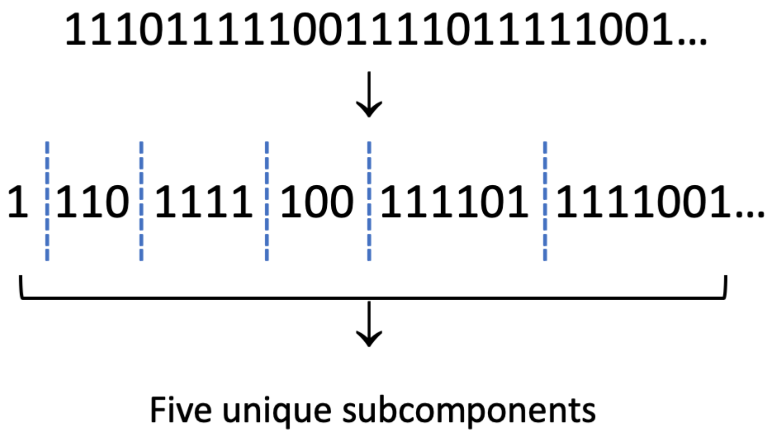

Figure 3. An example of how Lempel-Ziv takes a simple binary string of information, scans from left to right, and counts the number of unique subcomponents found. In this example, five unique subcomponents are identified in the binary string.

Schartner and colleagues (2015; 2017a; 2017b) sought to develop and optimise this LZc algorithm for use as a reliable measure indicative of global states of consciousness, across all variations of altered consciousness. They were specifically inspired by the success of the Lempel-Ziv compression algorithm in calculating the PCI metric (Casali et al., 2013) and endeavoured to develop a similarly useful measure of whole-brain EEG signal diversity, without the need for TMS. Previous applications of LZc to spontaneous EEG data have varied in the number and location of EEG channels recorded, often only measuring the temporal signal diversity from a single location on the scalp. To improve LZc computations, Schartner and colleagues (2015; 2017a; 2017b) have applied and adapted the LZc algorithm to be optimized for whole-head, multi-channel EEG recordings across time. Complexity models of consciousness suggest that the differentiated neural activity underlying conscious experience is not localised to a specific region (Sarà et al., 2011; Tononi, 2008), supporting this move from single-electrode to whole-head EEG. This version of LZc has been claimed to capture both temporal and spatial differentiation of the EEG signal across the whole head (Schartner at al., 2015).

To compute the LZc of EEG data (Figure 4), first, the multi-channel EEG recording is organised into a matrix, where each row represents a channel and each column represents observation in time. This data is converted back into a single binary string of data, then an LZc algorithm calculates the compressibility of the information by counting the number of unique binary 'words' in the data. For each segment of data, a single LZc value between 0-1 is generated, where 0 represents completely compressible, and 1 represents total randomness. For example, a LZc score of 0.7 reflects 70% total randomness on this normalized scale. Therefore, the more informational diversity in the multi-channel EEG recording, the more unique components will be present in the data, and the higher the LZc score produced (Schartner et al., 2015).

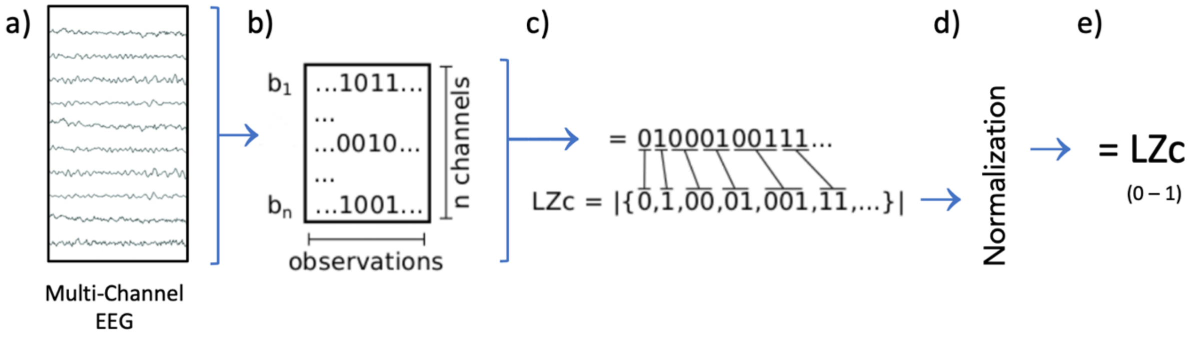

Figure 4. The computation of LZc for multi-channel EEG data. a) A segment of multi-channel EEG data is converted to binary and b) creates a matrix, in which each row is an EEG channel and each column an observation in time. This matrix is then c) converted to a single binary string where a LZc algorithm then counts the number of unique binary words in the data string. d) A normalisation step then divides this counted value by a randomized value to produce e) a score of LZc for that data segment (on a scale of 0-1) (Adapted from Schartner et al., 2017b).

In a set of three experiments, Schartner and Colleagues tested the relationship between LZc and consciousness, in different scenarios of altered consciousness: reduced consciousness during anaesthetic sedation (Schartner et at., 2015) and sleep stages (Schartner et al., 2017a), as well as altered diversity of conscious experience with psychedelic drugs (Schartner et al., 2017b). Results indicated that average LZc during Wakeful Rest and Mild Sedation was significantly higher compared to a loss-of-consciousness from Propofol anaesthesia. A small yet significant difference between Wakeful Rest and Mild Sedation (Schartner et al., 2015) was additionally displayed. Moreover, LZc for Wakeful Rest and rapid eye movement sleep (REM) was noted as significantly higher than during non-REM deep sleep. However, Wakeful Rest and REM presented no significant difference in average LZc (Schartner et al., 2017a). Interestingly, their most recent study found average LZc increased during psychedelic states induced by ketamine, lysergic acid diethylamide (LSD), and psilocybin. LZc during ketamine and LSD was significantly higher than during Placebo, but the difference between psilocybin and Placebo was small yet insignificant. A small correlation between LZc and subjective ratings of the intensity of psychedelic experience was also observed, but researchers acknowledged this requires further investigation (Schartner et al., 2017b). This increase in complexity is interpreted to reflect, for the first time, an increase in level of consciousness above the baseline of Conscious Wake (Schartner et al., 2017b).

This series of experiments by Schartner and Colleagues have been used to establish the LZc algorithm as a robust indicator of level of consciousness, whether due to physiological or pharmacological changes. In each experiment, a series of control tests were conducted, revealing that the main trend of results was largely unaffected by changes to parameters used in recording or analysis (e.g. segment duration, number of channels, location of channels, sampling rate). They also ran a series of tests to control for alternative sources of the observed effect (e.g. spectral power, influence of specific frequency bands, and correlations with other complexity algorithms). Across all tests, the main results were reasonably unaffected. They report finding “broadly identical results” and “no major changes” in the behaviour of LZc (Schartner et al., 2015; Schartner et al., 2017a; Schartner et al., 2017b). However, while the general trend of results may not be drastically altered, each change does modify the output of LZc, and the specific nature of these changes is not thoroughly investigated or reported upon. They posit LZc as a simple, practical measure, consistent with complexity theories of consciousness (Schartner et al., 2017b). However, a large portion of information is missing about how LZc behaves as a measure, regardless of the relationship to levels of consciousness. Exercising caution when making general claims about successfully discriminating between different levels of consciousness is essential. Given the significant clinical and ethical implications of using an inaccurate measure, these finer details should not be ignored.

Variability of LZc. The degree to which LZc of EEG may vary, either with or without a corresponding change in consciousness, is largely unknown. Currently, research is solely focused on working to explore the relationship between level of LZc and level of consciousness. This relationship is most frequently explored by contrasting a baseline conscious condition against an altered or reduced conscious condition. Therefore, a change in LZc up or down is concluded to reflect a change in consciousness up or down, relative to this baseline condition (e.g. Schartner et al., 2017b). However, this conclusion is necessarily reliant upon an assumption that Conscious Wakefulness is a relatively stable state and, consequently, that EEG complexity will be relatively stable during this period. If the LZc measure is unstable, then a small sample recording would not reflect the LZc across the whole period, irrespective of what causes variability. Therefore, there is an additional issue regarding practicality because it is assumed that a small sample recording will be representative of this period.

Recordings captured during conscious wakefulness are relied upon to be representative of this state and assumed to be stable enough to use as a baseline comparison condition. The baseline conditions used by Schartner and colleagues were different every time. They were always brief recordings (around 10-minutes of total data per condition) and not randomised to control for potential confounds (e.g. time of day). The condition itself also differed. For example, in the sleep and anaesthesia studies, participants were seated with eyes closed for a period of Wakeful Rest before the altered consciousness conditions (Schartner et al., 2015; Schartner et al., 2017a). In the psychedelics study, it was a different group of people in a placebo drug condition being compared against data from the active drug condition (Schartner et al., 2017b). In all three studies, only around 10-minutes of data remained per condition after preprocessing (Schartner et al., 2015; Schartner et al., 2017a; Schartner et al., 2017b). It is not typical to record for any longer than 30-minutes, but it is acknowledged that longer recordings, or repeated recordings across an extended duration of time, might prove to be beneficial (Sitt et al., 2014).

This is a typical representation of the short durations and varied conditions used as a baseline conscious wakefulness condition in this field of complexity measures of consciousness**.** However, these different conditions contribute to different results, which are not acknowledged or discussed in any meaningful way in the field. For example, Schartner and colleague’s (2015) study found no significant difference of LZc between Wakeful Rest and REM sleep, despite the clear difference in the global state of consciousness. Conversely, a similar study by Andrillon, Poulsen, Hansen, Léger and Kouider (2016) used an engaged cognitive task as their baseline Wakefulness condition, instead of an eyes-closed seated rest, and found there was a small but significant decrease in average LZc for REM sleep compared to Wake. It has additionally been presented that LZc is significantly lower when participants simply have their eyes closed, compared to eyes open (Farns, Jule, Nilsen, Romundstad & Storm, 2019). Irrespective of the empirical relationship between LZc and level of consciousness, to correctly interpret any measure as being increased or decreased from a baseline state we must first investigate and establish the scope of this baseline.

We cannot be confident that a change in LZc is associated with a change in the overall level of consciousness if it is unknown how LZc may vary without a corresponding change in consciousness. Although these studies display a change in complexity relating to a change in consciousness, no one has yet to chart how LZc behaves over an extended period of Conscious Wakefulness. Conscious Wakefulness is a long period of time that, although variable in experience across the day, typically does not involve any change in global state of consciousness. Tests for robustness and reliability of LZc have focussed on ruling out alternative explanations for the observed differences in LZc between different conditions of consciousness. Whilst there is seemingly compelling evidence for the reliability of LZc, there are many limitations that need to be acknowledged. Given that LZc is a measure of spontaneous EEG signal diversity, any factor that increases the regularity of signal or synchronicity between electrode channels will decrease the overall complexity (Schartner et al., 2015). However, if informational complexity measures of EEG are indeed reliable, robust and objective indicators of level of consciousness, they should also be relatively stable during this period. If not, this undermines the proposed purpose of the measure.

Due to the nature of this field's overarching clinical motivations, a thorough understanding of the variability of the measures being developed is crucial. If the claim is that a measure can indicate an individual's level of consciousness, the implication is that it may be applicable in clinical practice. Complexity during Wakefulness is assumed to be stable enough to consider any change from this baseline to be reflective of a change in global state of consciousness. However, it is possible that wakefulness is unstable and there may be periods where the complexity of the neural activity is reduced throughout the day at periods when the individual is clearly remaining conscious. This assumption of stability has never been empirically explored.

While the interests of using LZc in research are to differentiate between groups of different conscious conditions, these findings could lead to LZc being used to decide and define level of consciousness in clinical practice. The need for a clinically useful tool to improve diagnostic accuracy is significant, but there are also considerable ethical implications of using a tool we do not fully understand. In particular, the primary clinical motivation is to identify levels of overall consciousness in DOC patients who have reduced levels of consciousness that are relatively difficult to behaviourally identify (Schnakers, 2009). Using LZc, or any other EEG complexity measure, may contribute towards making significant ethical decisions about patient care. Some examples include identifying fluctuations in consciousness throughout the day, a diagnostic indicator of MCS (Owen, 2008), being used as evidence to determine whether behavioural signs are evidence of intentional or reflexive behaviour, or a to make the distinction between UWS and MCS (Giancino et al., 2004; Owen, 2008; Seel et al., 2010). These diagnoses determine conditions of ongoing patient care, pain management, and even life support decisions (Schnakers, 2009).

It is important we investigate and understand every aspect of how these complexity measures of consciousness behave, as well as which measurement parameters affect the calculation. At this point, information around these measures is insufficient. We do not know what encompasses LZc of the waking realm, or how it relates to cognition and behaviour. It is important to identify factors that need to be controlled or considered in future. To date, the stability of LZc during healthy wakeful consciousness has been assumed but has never been directly addressed. The claims and conclusions being made about LZc as a robust and reliable research measure, are reliant on this assumption at a foundational level. The empirical ability to discriminate between conscious and unconsciousness has led these measures to be suggested for clinical applications. However, it must be acknowledged that the current evidence cannot support this application. Therefore, it is unknown if LZc is a useful measure or if it can be used in the intended way.

Current Study

Aims

The overall aim of the current study is to provide a detailed investigation of the assumed stability of LZc over time. This novel insight will be achieved by utilizing the unique perspective of long-duration EEG recording. We will consider LZc across conscious wakefulness, a period in which individuals have no subjective change in their global state of consciousness. By analyzing the degree to which LZc varies across the waking day, this study will determine whether current methods of calculating a LZc score for healthy brain activity are likely to be providing an appropriate reference for clinical and practical uses. We will also provide a framework for assessing the reliability of LZc and promote the use of this framework to investigate all EEG-based complexity measures of consciousness. This study will discuss and address important, yet overlooked, assumptions in the field, significantly contributing to the growth and development of empirical measures of consciousness. Specifically, the key aims of this study are:

1. Replicate findings of differences in average LZc of sleep/wake states

Specifically, the first aim of this study is to replicate previous research, investigating differences in average LZc between different global states of consciousness. Our conditions will be conscious wakefulness (Wake) and Sleep (Stage 1, Stage 2, Slow-Wave-Sleep, and REM Sleep). In this study, the Wake condition will comprise of data from the full waking day, across multiple days, to provide a comprehensive and representative Wake condition. This will address the limitation of previous research using small and unrepresentative samples as their baseline conscious condition. It will also serve to validate the use of long-wear EEG, by replicating previous findings with this data.

2. Investigate the variability of LZc output during conscious wake

The second aim of this study is to address the assumption that LZc is stable when consciousness is stable. We will do so by investigating the output of LZc across Conscious Wakefulness; a period where it is assumed that people are fully conscious in some relatively stable way, with no change in global state of consciousness. This will be a thorough, descriptive insight into a very broad, all-encompassing conscious condition across an extended duration of time (up to 14-hours per day per subject). This will empirically establish the presence or absence of variability in LZc.

3. Examine the impact of LZc variability on the reliability of samples

The third aim of this study is to investigate whether Wake is a condition that is likely to be an appropriate reference or baseline condition for research and clinical practice. This will establish the practical impact of any observed variability in LZc output. To test this, we will take small duration samples, around the size typically used in research, and calculate all possible mean LZc scores that can be obtained from a sample that size. This will quantify the impact of any variability on the practical usage of this measure.

4. Promote long-duration EEG as an investigative framework

Finally, we aim to use this study to promote the use of long-duration EEG as a framework by which to assess all EEG-based complexity measures of consciousness. This methodological approach provides a unique and comprehensive insight into complexity across extended durations of time, including different global states of consciousness and the transitions between them. We hope to redirect discussion in this field towards key underlying issues and assumptions, while also promoting a framework to empirically address them.

Method

Participants

Long-duration continuous EEG data from two participants was used in this study. Both participants provided informed consent for their data to be used for scientific research. No compensation was given for participation. The use of this data for research was approved by St Vincent's Human Research Ethics Committee (reference number LNR/16/SVHM/243).

Participant one (P1) was a 74-year-old female who was referred to St Vincent's Hospital for long-duration EEG recording for neurological diagnostic purposes. It was confirmed, from this multi-day recording period, that this participant did not have neurological epilepsy. They did, however, experience a psychogenic non-epileptic seizure (PNES) event during the recording. This PNES event was exhibited as a reported loss of time by the participant but did not display abnormal electrical activity in the EEG recording. All data from this period was excluded from further analysis as the PNES event was not of interest to this study. Participant two (P2) was a healthy 40-year-old female, recruited via word of mouth as an additional healthy control dataset for this study. P2 had no history of epilepsy, migraine, or neurological injury.

Both participants underwent a continuous, multi-day recording period using a portable EEG recording system. After being fitted with the EEG equipment, they were sent home for the duration of the study. Participants were free to move about and were instructed to continue with their daily lives. Caffeine, nicotine, and alcohol consumption was monitored but not restricted. Both participants were on no medication. No behavioural manipulations or tasks were conducted.

Data Acquisition

Participants were fitted with an electroencephalogram-electrocardiogram (EEG-ECG) recording system by a trained neurophysiologist at the Seer Medical research office. Data was recorded using a portable Compumedics Siesta amplifier and data acquisition system (Compumedics Limited). This portable system recorded 22-channels of neural and cardiac electrical activity. 21 silver electrodes (Compumedics Limited) were attached to the scalp in a standard 10-20 montage (see Figure 5) (Jasper, 1958) using glue (SLE collodion adhesive). Hair was parted aside to attach the electrodes and a conductive gel (Ten20 Conductive Paste, Weaver and Company) was used between the scalp and electrodes. The ECG electrode was attached to the chest, using a commercially available stick-on electrode (Red Dot, 3M) and medical tape (Fixomull stretch). This ECG electrode was used to record heart rate activity.

The data was recorded at a sampling rate of 256 Hertz (Hz) referenced to the central electrode (Cz) and input impedance for all electrodes was automatically kept below 10 kilo-Ohms (kΩ). Wires connected the electrodes to a portable recording and transmission system. This was a small system, which the participants wore on a belt around their waist. This belt also contained a battery pack that was changed daily (refer to Figure 6). The data was recorded in real-time and transmitted wirelessly via Wi-Fi to a nearby host server computer (Compumedics PDF). Additionally, all data was streamed via Wi-Fi to the Seer Medical cloud network. All data was synced to local Australian Eastern Standard Time (AEST).

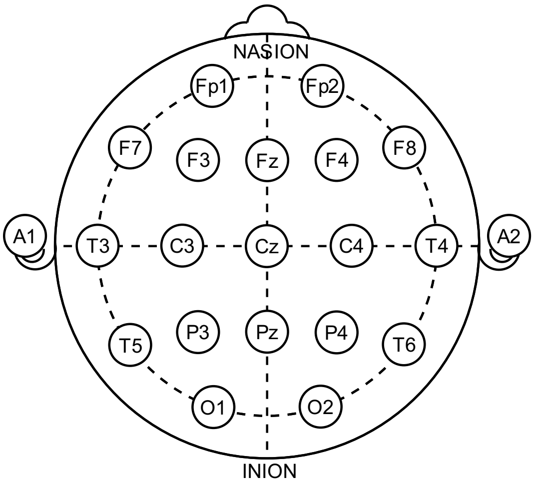

Figure 5. An example of a 10-20 montage. Image shows the topographical locations of electrodes on the scalp. 21 EEG scalp electrodes, including two mastoid electrodes for reference (A1, A2). All electrodes referenced to the central electrode (Cz) for recording. This montage is commonly used in EEG research, particularly in sleep studies, as well as epilepsy monitoring and long-duration EEG studies (From Mumtaz & Malik, 2018).

Video footage was captured during the recording period using a portable camera set-up. This was a wide-angle camera that could record the view of a whole room. It was connected to the same wireless computer system that was recording the EEG-ECG data. The video was recorded at 30 frames-per-second and was synced to the EEG-ECG recording file in AEST. Audio was also recorded by the camera set up. This camera was positioned on a stand facing the participant and was moved around by the participant during the day, to ensure they remained in view. The camera recorded video of the participants during sleep to assist in sleep scoring of the EEG data.

This portable long-wear EEG system recorded continuously for the duration of each participant's study. P1's wireless computer system and video camera were set up and only used in their home. Since the system only has a wireless recording range of 30-meters, EEG-ECG data intermittently stopped recording for periods during the day while the participant was out of the house. This is reflected by time periods of no-data during the recording, which were subsequently excluded from further analysis. P2 was a healthy control recruited for the purpose of this experiment. Therefore, they were instructed to carry the portable computer system around with them in a briefcase carry bag to avoid this issue of signal drop-out. EEG-ECG data was recorded continuously for the whole duration of their study. P1's video was only recorded intermittently during the day during periods where they were stationary and able to set it up to record. However, video was continually recorded at night to assist with sleep scoring of the EEG data. Global Positioning System (GPS) data for P2 was collected via a smartphone using Google Maps timeline. This information was used to inform researchers of the participant's activities during periods of missing video data. At the end of the recording duration, participants returned to have the electrodes and equipment removed by the trained neurophysiologist.

Figure 6. Example images of the portable EEG-ECG recording system in practice. The system is first attached by a trained neurophysiologist at Seer Medical (top). Electrodes are glued to the scalp. Wires are connected, gathered, and brought down the back of the participant's head. These are held in place with enough slack for the participant to move their head and neck without pulling at the wires. These wires all connected to the recording system, transmitter, and battery pack worn by the participant (middle). Participants are then sent home to continue about their daily lives (bottom). They are free to move about and complete daily tasks with little restriction from the equipment (From https://www.seermedical.com).

EEG Data Preprocessing

The data from each participant was preprocessed and analysed independently. The continuous EEG data for each participant was segmented into consecutive, non-overlapping, 10-second segments for processing and analysis. The full duration of recorded data was analysed from start to finish of the recording period. First, the raw EEG data from the Compumedics software recording was converted to Matlab-compatible files in MATLAB R2015a (a 32-bit version of MATLAB compatible with the conversion script used). EEG data was then analysed offline in MATLAB R2017b (64-bit). The ECG heart rate electrode channel, as well as the A1 and A2 mastoid electrodes channels, were removed from the analysis, leaving 19 scalp electrode channels. A Cleanline toolbox applied a notch filter at 50 Hz and 100 Hz to remove electricity mains artefacts from the data. A Current Source Density toolbox was used to apply a surface Laplacian Filter, a method of spatial filtering used to decorrelate adjacent electrodes. This was done to localise the source of activity more focally to each electrode. Linear detrending and baseline subtraction were then applied to each channel of each segment. This was followed by a Hilbert transformation, used to assess time-frequency information and prepare the data to be converted to binary information for the LZc analysis.

LZc Analysis

The preprocessed data was analysed with a published and available LZc algorithm sourced from Schartner and colleagues (2015). This algorithm is specifically designed to compute the LZc of multi-channel EEG data across-time. This LZc algorithm quantifies the diversity of information observed in the EEG data segment, proportional to the maximal amount of informational diversity possible (Schartner et al., 2017a). A single LZc value is produced for every consecutive data segment. A segment size of 10-seconds in length was selected based on previous research (Schartner et al., 2015; Schartner et al., 2017a; Schartner et al., 2017b).

The LZc compression algorithm counts the number of unique subsequences in the binary string. These unique subsequences are referred to as words. If a repeated word is identified, it is only stored once; therefore, the total data is compressed and there is no loss of information. Each new binary subsequence found is stored as a new word. The LZc algorithm creates a new dictionary of binary word strings as it computes each data segment. A normalisation step then divides this counted value by a value obtained from randomly sequence shuffling the same fixed-length segment of data. This shuffled segment is completely random and therefore represents maximum LZc for that segment (Schartner et al., 2015). By dividing the counted LZc value by the sequence shuffled value, the counted LZc of each segment can be presented as proportional to maximal complexity. LZc for all segments is therefore represented on a scale from 0 to 1, where 0 is completely compressible and 1 is completely random. The more diverse the data, the more unique components, and the higher the LZc score will be. For every segment of EEG data, the LZc algorithm computes and outputs a single LZc value.

To compute the LZc of multi-channel EEG (see Figure 7), each 10-second segment of preprocessed EEG data is first input, and each channel of data is converted to binary information. This multi-channel time series of binary EEG data is organised into a matrix. In this matrix, each row represents a channel and each column represents an observation in time. This multidimensional matrix is then converted, observation-by-observation, into a single binary string. A standard open-source LZc algorithm (Rosetta Code) searches for the number of unique words in the binary string and counts them. A normalisation step divides this counted LZc value by the sequence shuffled value, giving an output of LZc for each segment on a scale of 0-1. If there are any segments containing missing data, the LZc calculation fails, and the value output for that segment is 1. These segments containing missing data are removed from further analysis. These LZc calculation steps are repeated for every 10-second segment of data.

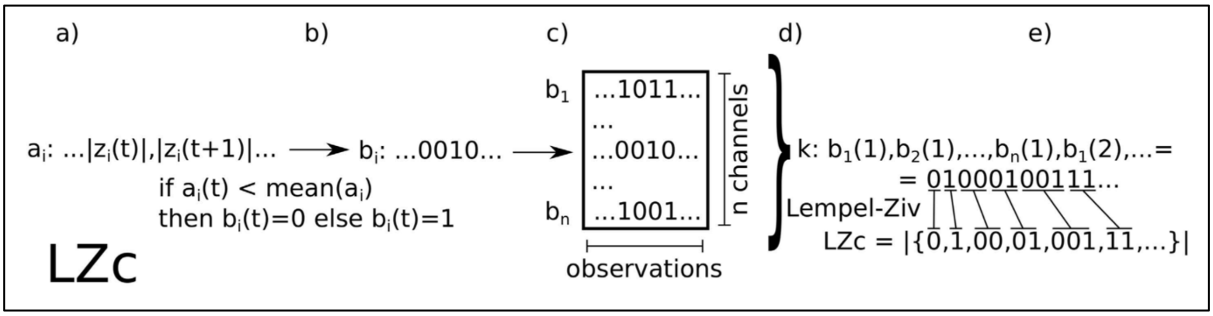

Figure 7. A schematic illustrating the computation process of the LZc algorithm. a) A 10-second segment of multi-channel EEG is converted to b) binary information. These strings of binary data are then used to create c) a matrix where each row is a channel and each column is an observation across time. This is then converted d) into a single binary string, observation by observation, and then e) the Lempel-Ziv algorithm counts this binary string for the number of unique binary words (From Schartner et al., 2017b).

Sleep Scoring

Sleep scoring was performed on the EEG data using Profusion PSG 3 software by an experienced sleep researcher using standard clinical criteria adapted from the American Academy of Sleep Medicine (AASM) criteria (AASM, 2007). The same researcher scored both datasets. Electromyogram (EMG) facial muscle activity artefacts and electrooculogram (EOG) eye movement activity artefacts were not quantified. Therefore, eye movements were assessed using Fp1 or Fp2 electrodes. This is a limitation that may result in an underestimation of total REM sleep, as well as limiting identification of arousals during REM. Electrode channels were re-referenced to the contralateral mastoid electrode. Video recordings were used to confirm the presence of arousals or movement of the participant, supplementing the sleep scoring process. Arousals during sleep were identified and labelled accordingly. A sleep score was provided for every 30-second segment of EEG data. The categories given were Wake, Stage N1 Sleep (S1), Stage N2 Sleep (S2), Slow-Wave-Sleep (SWS), REM Sleep, and periods of Wake-After-Sleep-Onset (WASO). The category of Wake was further defined based on time of day, to separate the 30-minutes post-sleep after waking up in the morning and the 30-minutes pre-sleep before falling asleep in the evening. Hence, Wake was defined as all data recorded during the waking day, excluding these pre- and post-sleep periods.

Category labels of sleep stages were then lined up with the LZc output along a timeline, for the duration of each participant's study. Sleep scoring was completed on 30-second segments; however, LZc scoring was completed on 10-second segments. Therefore, all three 10-second segments contained within a single 30-second block were all labelled with the corresponding sleep score for that block. These categories were used to group LZc data for analysis and form decisions regarding segment exclusions. Segments containing arousal from sleep (3- to 15-second long microarousals) were excluded from analysis. Periods of arousal longer than 15-seconds (categorised as WASO) were also excluded. During sleep stages, only periods of data that consisted of a 2-minute or longer consistent stretch of a single sleep stage were retained for analysis. This disregarded periods of fluctuation between sleep stages and limits inclusion to data from periods of consistent sleep stages. The pre- and post-sleep 30-minute blocks were eliminated from further analysis, as these sleep-wake transition times were not of interest to this study. Any segments identified as No Data (either identified during the sleep scoring process or segments with a LZc score of 1 due to missing data) were removed, as well as all data recorded during the equipment set-up stages at the beginning and end of the recording. The PNES event in the data from P1 was also excluded as this was not of interest to this study. After exclusions, 112-hours and 47-minutes of data remained for P1, and 65-hours and 35-minutes of data remained for P2.

Results

SPSS (Version 25) was used for initial data cleaning, coding of variables, computation of variables, assumption checking, and analyses. Descriptive analyses identified potential outliers, out of range values, and missing data values. No outliers, out of range values and missing data values were found. An alpha level of .05 was used for the significance threshold of all subsequent analyses.

LZc of Sleep and Wake Categories

An analysis of the LZc of Sleep/Wake categories of each participant replicated the findings of previous research. Specifically, the categories considered include: (1) Wake; (2) S1 Sleep; (3) S2 Sleep; (4) SWS; and (5) REM Sleep. As shown in Table 2, mean LZc was highest during Wake and lowest during SWS for both participants. As shown in Figure 8, the pattern of LZc for each Sleep/Wake category was similar for both participants.

Table 2.

Descriptive statistics of LZc in each Sleep/Wake category

	Participant 1			Participant 2		
Sleep/Wake Stage	*n*	*M*	*SD*	*n*	*M*	*SD*
Wake	26550	.67	.1	16244	.64	.08
S1 Sleep	432	.62	.12	120	.53	.11
S2 Sleep	7982	.57	.11	4404	.5	.11
SWS	1740	.52	.04	1149	.45	.05
REM Sleep	3900	.57	.14	1692	.51	.1

Additionally, Figure 8 shows considerable variation across all Sleep/Wake categories as demonstrated by the error bars, which indicate one standard deviation (SD) above and below the mean. A coefficient of variance (CoV; a measure of SD relative to the mean) was calculated for each Sleep/Wake category (P1 and P2 separately) to compare variation across categories. The CoV calculates SD relative to the mean of each category, representing variability as a relative percentage. This revealed that within-category variation was lowest in SWS for both participants (P1 8.34%, P2 10.13%). Additionally, variation was highest in REM Sleep for Participant 1 (24.97%) but for Participant 2, variation was highest in Stage 1 (21.40%) and Stage 2 (21.04%). All CoV values are presented in Appendix A.

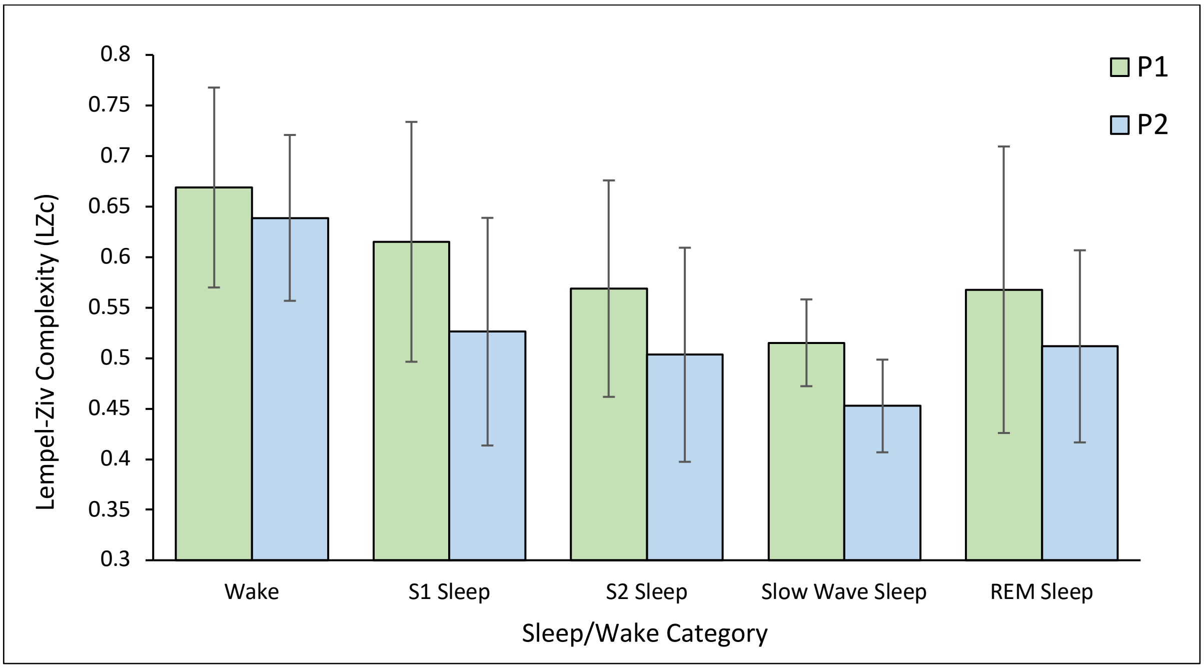

Figure 8. Clustered bar chart of mean LZc values across Sleep/Wake categories. Error bars reflect one standard deviation above and below the mean.

P1. A one-way Analysis of Variance (ANOVA) was used to examine whether, for P1, LZc significantly differed between Sleep/Wake categories. A *Levene's F* test revealed that the assumption of homogeneity of variance was not met. Therefore, *Welch's F* test was used.

This revealed a significant difference in mean LZc across Sleep/Wake categories for P1, *Welch's* $F(8, 2365.75) = 2473.62,\ p <.001$. It can be concluded from this test that at least two of the five Sleep/Wake categories significantly differ in average LZc scores. A Games-Howell post-hoc comparison was used to determine between which groups mean LZc significantly differed. These results indicate mean LZc to be significantly higher during Wake (p < .001, *Cohen's d* > .5), compared to all other Sleep/Wake categories and found mean LZc to be significantly lower during SWS (p < .001, *Cohen's d* > .7), compared to all other Sleep/Wake categories. Moreover, S1 Sleep was significantly higher than S2 Sleep (p <.001, *Cohen's d* > .4) and REM Sleep (p < .001, *Cohen's d* > .3). However, mean LZc was not significantly different between S2 Sleep and REM Sleep (p > 1).

P2. A one-way ANOVA was used to examine whether, for P2, LZc significantly differed between Sleep/Wake categories. A *Levene's F* test revealed that the assumption of homogeneity of variance was not met. Therefore, *Welch's F* test was used. This revealed a significant difference in mean LZc across all Sleep/Wake categories for P2, *Welch's* $F(7{,}1276.76) = 2898.01, p <.001$. It can be concluded from this test that at least two of the five Sleep/Wake categories significantly differ in average LZc scores. A Games-Howell post-hoc comparison was used to determine between which groups mean LZc significantly differed. These results indicate mean LZc during Wake to be significantly higher (p < .001, *Cohen's d* > 1.1) than all other categories, and mean LZc during SWS to be significantly lower (p < .001, *Cohen's d* > .6) than all other categories. S1 Sleep, S2 Sleep, and REM Sleep did not significantly differ from each other (p > .05).

Variability of LZc During Conscious Wakefulness

An examination of the LZc data during Wake was undertaken to investigate variability. This is a period of relatively stable consciousness, therefore, any variability in LZc cannot be explained by changes in the global state of consciousness. For each participant, Wake consisted of up to 14-hours of consistent, long-duration EEG recording per day, from 30-minutes after waking up to 30-minutes before going to sleep. Data was collected across multiple days. Descriptive statistics for each day of Wake data is presented in Table 3. In total there was 74-hours 11-minutes 20-seconds of Wake data used for Participant 1, and 45-hours 7-minutes 10-seconds of Wake data used for Participant 2.

Table 3.

Descriptive Statistics of LZc for each day of Wake data

	Participant 1			Participant 2		
Day of Wake	*n*	*M*	*SD*	*n*	*M*	*SD*
Day 1	1314	.65	.08	4159	.65	.07
Day 2	3594	.68	.07	5169	.63	.08
Day 3	4514	.67	.07	4965	.63	.09
Day 4	4107	.66	.09	1950	.64	.09
Day 5	4406	.67	.1	-	-	-
Day 6	3416	.7	.07	-	-	-
Day 7	4301	.65	.15	-	-	-
Day 8	1056	.7	.11	-	-	-
Total	26708	.67	.1	16243	.64	.08

LZc across time was plotted for each participant, as shown in Figures 10 and 11. Each line represents a 5-minute moving average of LZc data across one day, with all days overlaid. As shown, there is slight variation throughout the day. However, the overall range remains relatively stable. There is a visible feature of occasional sharp low peaks of LZc, in both participants' data. Despite the data being smoothed with a moving average line, these 5-minute averages reach as low as .19 for P1 and .27 for P2. The range between the highest and lowest moving average value is .61 and .51 respectively, and the interquartile ranges are .06 and .05.

The raw LZc data was then further expanded and visualized, also presented in Figures 10 and 11. These figures show a single LZc data point for every non-overlapping 10-second segment across-time during Wake. This allowed a non-smoothed, comprehensive view of the LZc data as it looks across-time. There exists a considerable, but consistent, scattering of LZc data points below the mean, for both participant datasets. This feature of the data is reflected in the percentage frequency distributions of Wake data as shown in Figure 9, where both participant datasets show a tail of lower LZc scores. This is additionally reflected by a negative skew of LZc during Wake for both participants. For P1, this was a highly negative skew (skewness = -2.46), and for P2, this was a moderate negative skew (skewness = -.95).

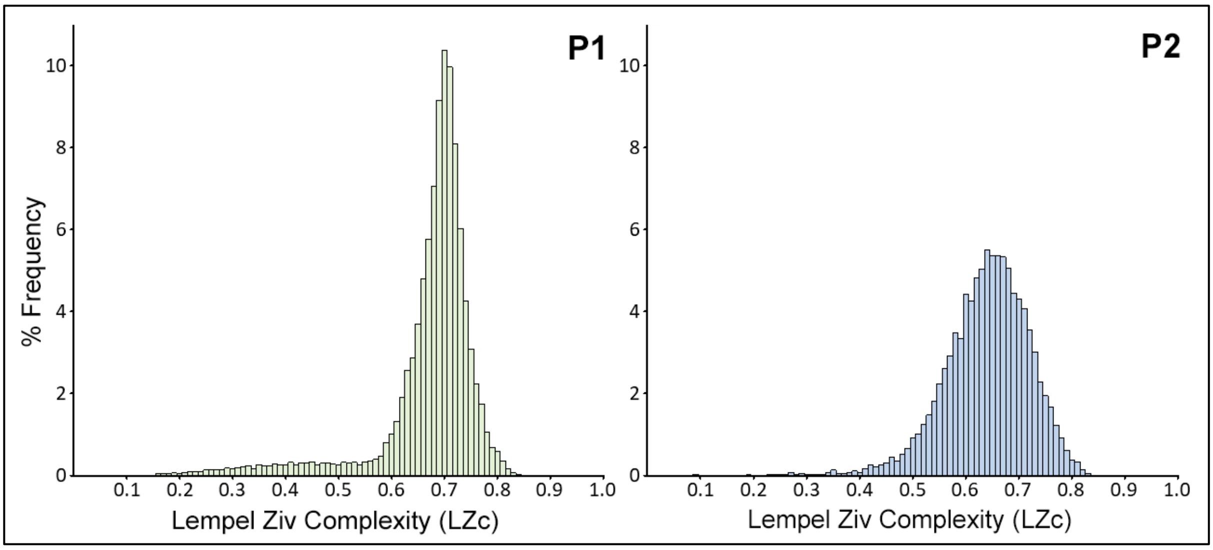

Figure 9. Histograms showing the percentage frequency of LZc scores during Wake. Participant 1 (left) and Participant 2 (right) shown separately.

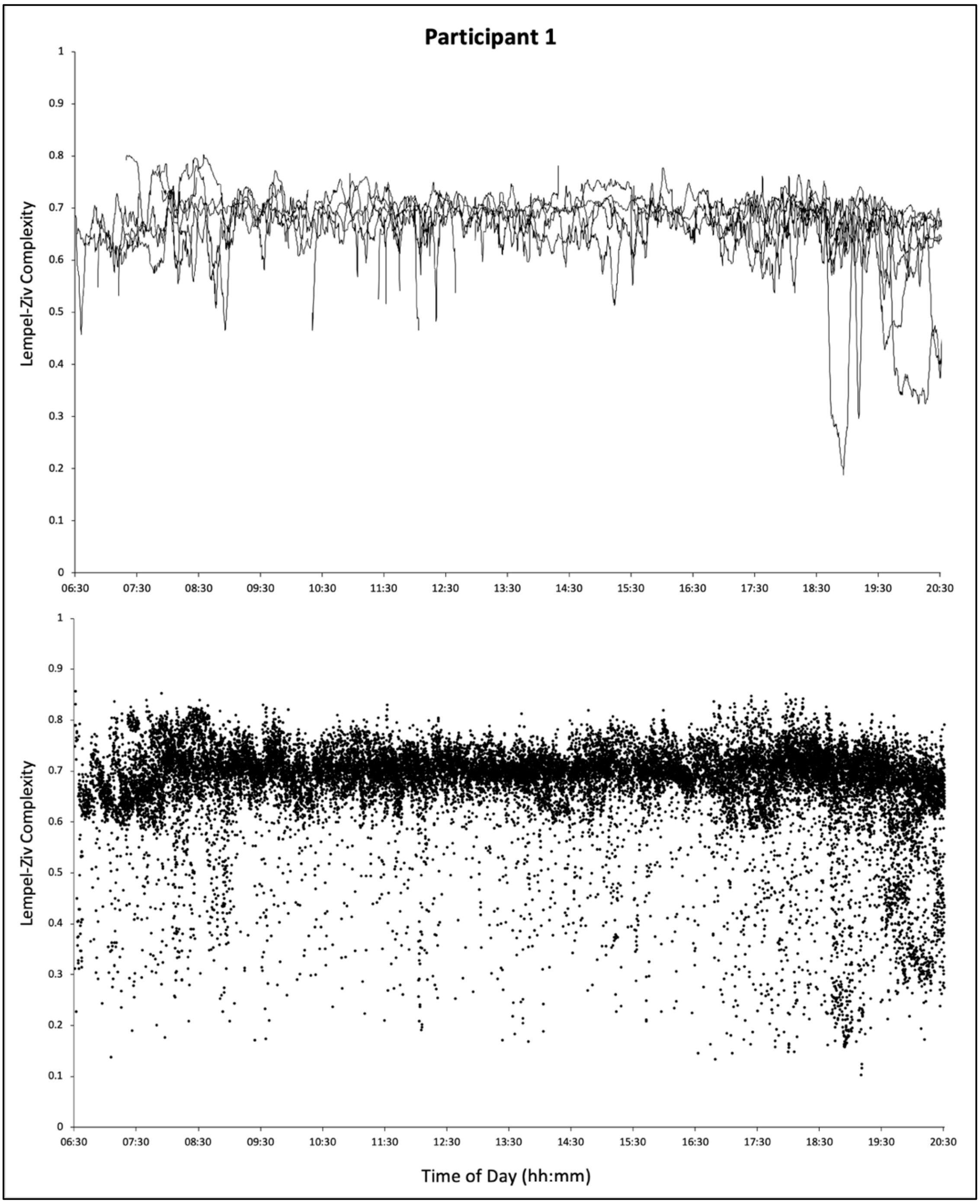

Figure 10. Graphs showing LZc across-time for Participant 1 during Wake. This 14-hour period is relative to the individual's sleep schedule during the recording period and therefore, Participant 1's data is shown from 6.30 am to 8.30 pm. 8-days of data are overlaid onto the same axis. Periods of No Data are reflected by gaps in the data lines on the graph. The first graph (above) shows a smoothed representation of LZc across Wake, for each day. Each line reflects a five-minute moving average of the data. The second graph (below) is a non-smoothed scatter graph with the same data on the same axes, showing the raw LZc output, with a single LZc value produced every 10-seconds.

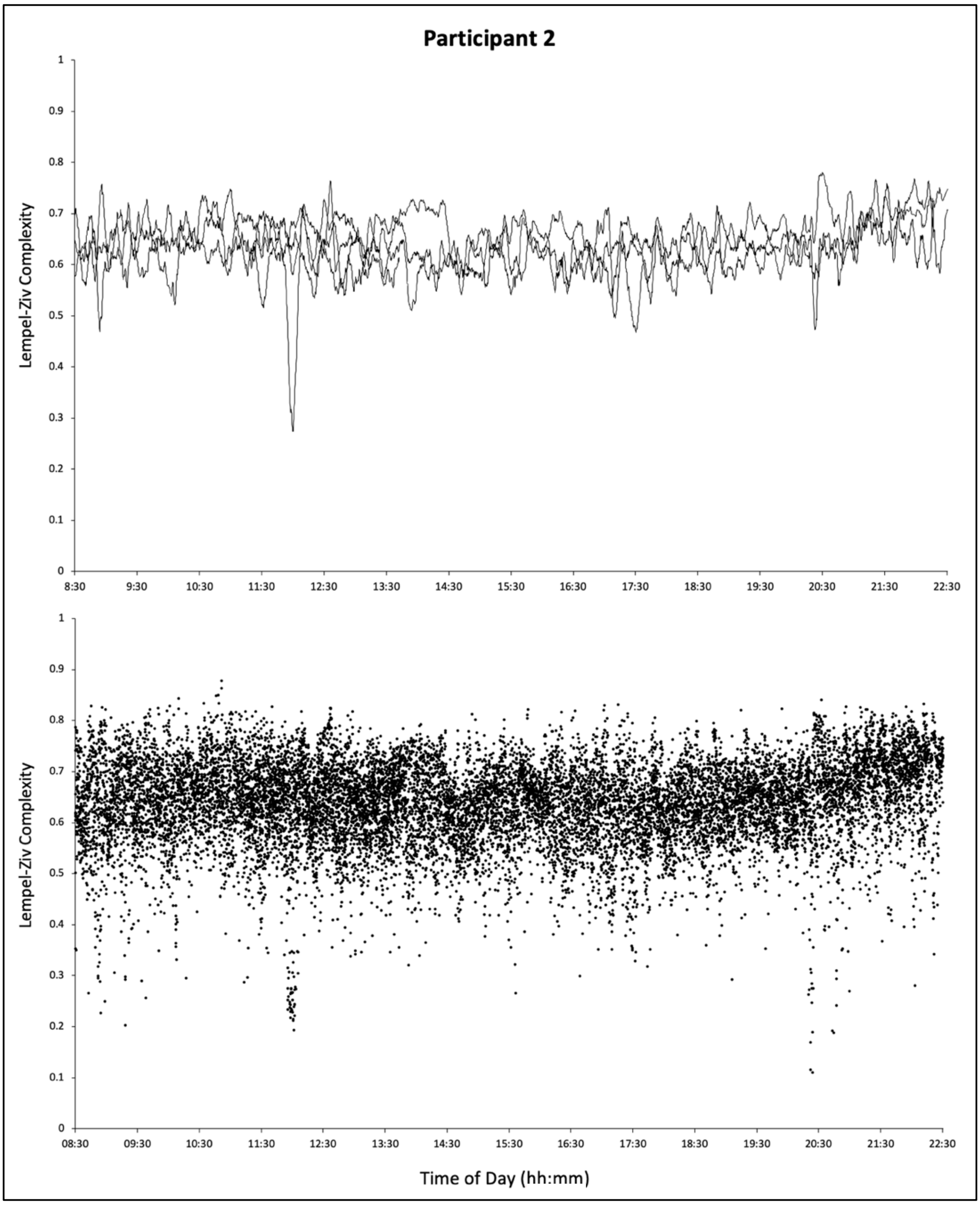

Figure 11. Graphs showing LZc across-time for Participant 2 during Wake. This 14-hour period is relative to the individual's sleep schedule during the recording period and therefore, Participant 2's data is shown from 8.30 am to 10.30 pm. 4-days of data are overlaid onto the same axis. The first graph (above) shows a smoothed representation of LZc across Wake, for each day. Each line reflects a five-minute moving average of the data. The second graph (below) is a non-smoothed scatter graph with the same data on the same axes, showing the raw LZc output, with a single LZc value produced every 10-seconds.

LZc Sample Reliability

Given the variability of LZc data during wakefulness, we investigated the impact this has on the variability of mean LZc values, obtained via small samples. These small samples represent how LZc is typically used in research and therefore, contributes directly to understanding the use of LZc in practice. To create the LZc samples, overlapping moving averages were created using Wake data for each participant. This was repeated for a series of different sample durations. A comprehensive set of all possible mean LZc values for each sample duration was created, across a range of durations. A box-plot for a subset of these sample durations is shown in Figure 12. This figure shows that the median LZc value obtained from samples is consistently lower than the overall median value for all Wake data, across all sample durations, for both participants.

A CoV was calculated for each sample duration group, as a way to compare relative variability. The CoV value for each sample duration group is provided in units of standard deviation as a percentage of the mean for that group. This reflected the degree of variability of LZc sample means, at each sample duration. As shown in Figure 13, the CoV decreased as the sample size duration increased. Both participants showed a similar pattern. While this pattern of decreasing variability with increasing sample size is not surprising, it is informative to understand how this variability relates to sample size duration in minutes of EEG recording. Additional CoV values are presented in Appendix A, alongside other CoV that were calculated.

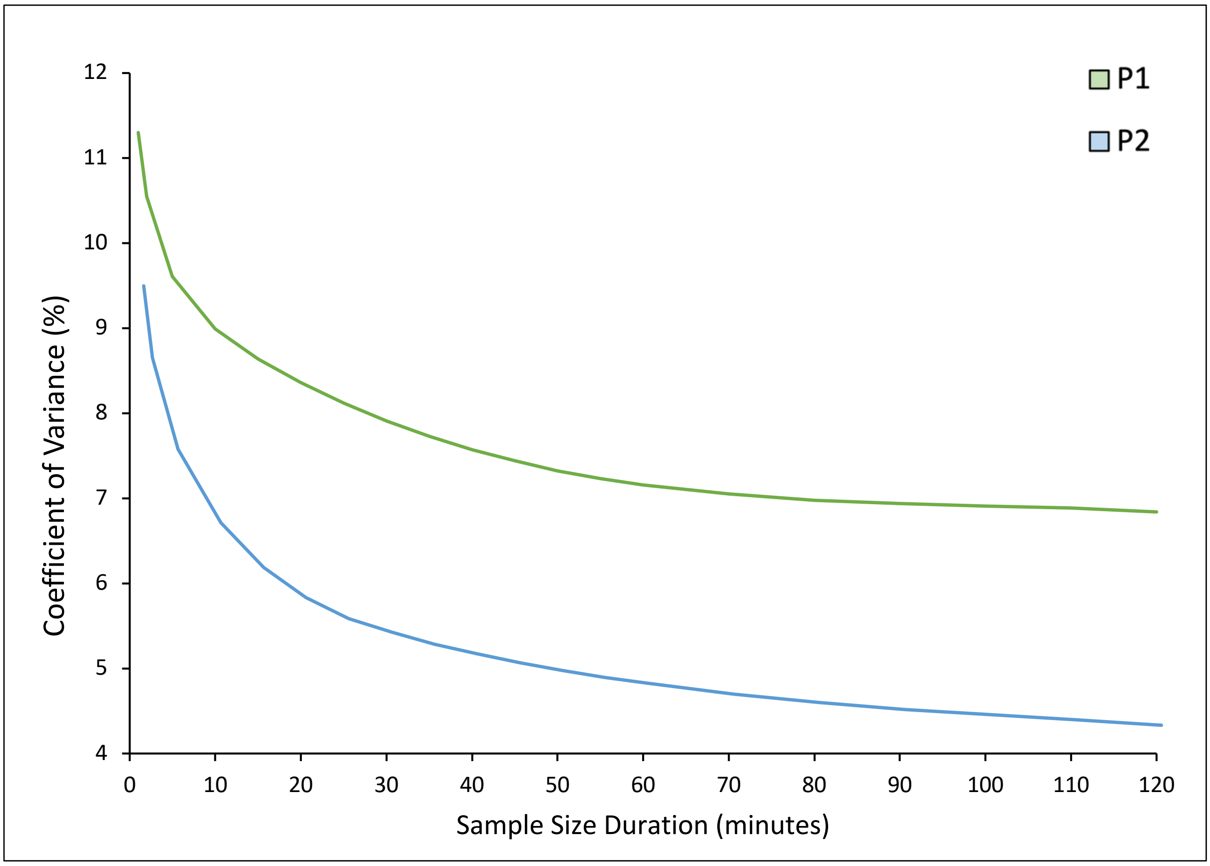

Figure 12. Coefficient of Variance at each sample size duration for Participant 1 and Participant 2 (overlaid onto the same graph).

To test whether there was any stability of LZc at a larger sample duration, the LZc data during Wake was divided into separate days, based on the day the data was recorded (for example, Day 2 of Wake data for P1 includes all waking data on the second day of recording, from 30-minutes after waking up to 30-minutes before falling asleep). A CoV for each day was calculated (see Appendix A), with Participant 1 showing a notably higher variation on Day 7 (23.71%) compared to all other days, with CoV ranging from 9.70%-15.78%. For Participant 2, variation was relatively more consistent across days, with CoV ranging from 11.27%-13.57%.

A one-way ANOVA was used to determine if average LZc during Wake significantly differed between days. Each participant's data was analysed separately. Levene's test for homogeneity of variance was not met for either dataset [P1 ($F(7,26700) = 495.35$, $p < .001$);

P2 (*F* (3,16239) = 12.26, *p* < .001)], meaning the assumption of homogeneity of variance was unfulfilled. Therefore, Welch's adjusted F ratio investigated significant differences between groups. This test was significant for both P1 (*F*(7,7890.2) = 96.9, *p* <.001) and P2 (*F*(3,6913.18) = 83.49, *p* <.001), indicating a difference between two or more of the different days of recording. Therefore, even when considering the average of a sample up to 14-hours in duration, one sample within a person during conscious Wake can be significantly different from another sample.

Discussion

Overview

This study identified variability in LZc with no corresponding change in consciousness. LZc is a complexity-based measure of spontaneous EEG signal diversity that has previously been proposed to successfully index level of consciousness in an individual (Andrillon et al., 2016; Ferenets, Vanluchene, Lipping, Heyse, Struys, 2007; Schartner et al., 2015; Schartner et al., 2017a; Schartner et al., 2017b; Sitt et al., 2014). Previous research empirically established a relationship between LZc and different global states of consciousness by comparing the LZc during a baseline conscious condition against an altered or reduced conscious condition. However, these claims rely on the assumption that if consciousness is stable, LZc is also stable. This has never previously been empirically tested.

In this study, we address this assumption by utilising long-wear EEG to comprehensively study the output of this measure across multiple days and nights. Specifically, we first aimed to replicate previous research by comparing LZc during a conscious state (wakefulness) against LZc during a reduced conscious state (sleep), with the addition of using comprehensive recording periods from each state instead of small sample recordings. Next, the LZc across conscious wakefulness was explored in more depth, to assess whether the output of this measure was stable during this state of relatively stable consciousness. Additionally, we investigated the effect of this variability on the practical use of the measure by testing the range of possible sample means that could be identified during conscious wakefulness. Lastly, we showed that this investigative framework, utilising long-duration EEG, provides an information-rich dataset, and succeeded at identifying variability in LZc across time.

Results showed that, on average, LZc during conscious wakefulness was significantly higher than during all stages of sleep, for both participants. This finding is consistent with the trend of results found in previous research, suggesting that this effect prevails even when using a comprehensive recording of Wake as a baseline condition. This also serves to validate the use of long-duration EEG, as it produced results comparable to other EEG recording equipment. Upon further investigation into LZc during Wake, widespread variability of LZc around the mean was revealed. This spread of variation was relatively consistent across time, although the distribution of LZc presented a negative skew. This skew was reflected in the raw LZc data, where there was a visible presence of scattered low LZc values across the day, and in the percentage frequencies, where there was a tail of low LZc values. Data from both participants showed the same pattern of results, although sitting at different median values on the scale of LZc. When samples (ranging from 1- to 120-minutes in duration) are taken from this variable data, the resulting sample means are also highly variable. The smaller the sample duration, the more extensive the range of possible mean LZc scores which could be obtained. It was even found that mean LZc within a person over one day (averaged over up to 14-hours of data) could significantly differ from the next.

The findings of this study provide valuable insights into the variability of LZc as well as prompt critical discussions regarding complexity-based measures that are claimed to indicate level of consciousness. Our results reveal that there is a wide range of variation in the LZc of EEG signal across conscious wakefulness and that this variation impacts the mean LZc obtained from smaller duration samples as well. This variation raises questions about the practicality of the typical implementation of this measure in research. As it is currently used baseline control conditions upon which all findings are compared against is unreliable. LZc is not stable during conscious wakefulness, a period consisting of no change in global state of consciousness, as previous research has assumed. These results directly address the assumption of stability, which previous research has relied upon, and fills this gap in knowledge. These results cast doubt on the reliability of the measure and its use in research as well as future use in clinical practice.

There is a great, clinical urgency for an objective, reliable, brain-based measure that can provide insight into overall levels of consciousness. Currently, clinical practice is largely reliant on behavioural observations and physiological measures to indicate consciousness, which are known to be limited (Bender et al., 2015; Giancino et al., 2014; Schnakers, 2009). Additionally, in any state of reduced or altered consciousness, interaction with the environment is likewise reduced or altered (Koch et al., 2016; Sanders et al., 2012). Therefore, we have limited insight into these experiences and thus, a limited ability to study them. Complexity-based measures of spontaneous EEG data have recently gained traction as a tool for this purpose. Suitably, they are non-invasive, cost-effective, do not require active participation from the participant, and provide an overall metric of global state of consciousness (Blume et al., 2015). Of these complexity measures, LZc has received a lot of focus in recent research but there are other examples that are similar in nature. While the need to develop a tool that improves clinical diagnoses is imperative, it is more important to ensure that any implemented measures are accurate and reliable before they are used to produce claims regarding a patient's level of consciousness. These measures must be thoroughly researched and prove their practical utility for the intended applications. This is critical information to establish since there are serious ethical implications for not doing so.

It is concerning to see changes in LZc without any corresponding change in consciousness. This measure is promoted as a "reliable index of level of consciousness" (Schartner et al., 2017b), and is being used with the intention to infer knowledge about the level of consciousness within an individual. Additionally, it is marketed towards future clinical diagnostic purposes. The push to establish LZc and other complexity measures, as robust, reliable, indicators of consciousness neglects to acknowledge that there are several essential issues to firstly address. Whilst the relationship between complexity and consciousness is being empirically established, we must also empirically address the assumptions. The variability identified in this study is concerning, irrespective of the cause, and must be understood in order to eventually understand how LZc relates to consciousness.

The results of this study directly impact our understanding of previous LZc research and inform the direction of future research. With this paper, we hope to draw the focus of discussions in this field to these important issues. These are not only issues for LZc, but for all complexity-based measures of spontaneous EEG. We also discuss the methodological approach of using long-wear EEG to address these issues directly and encourage the importance of completing feasibility studies, such as these, on all current and future complexity measures. The purpose is not to rule out LZc as a feasible measure indicative of consciousness but to show that feasibility studies, as exemplified here, are a useful and successful way of identifying variability in complexity measures of consciousness.

Discussion of LZc Results

LZc of wake is significantly higher than during sleep. The first finding of this study replicated that of previous research, looking at significant differences in average LZc between different global states of consciousness. This was important to replicate with our data due to the choice of Wake condition we used, as well as the long-duration EEG equipment. Instead of utilising a brief recording from Wake or Wakeful Rest, we used data from this whole range, combined across multiple days for each participant. The Wake condition used was the average of the entire comprehensive recording. This is a more representative measurement of conscious wakefulness and thus, a more reliable baseline condition, compared to what is typically used in LZc research. The reduced consciousness conditions in this study were stages of natural sleep, across multiple nights: S1 Sleep, S2 Sleep, SWS, and REM Sleep. Results showed that, for both participants, average LZc during Wake was significantly higher than all sleep stages, and average LZc during SWS was significantly lower than all other sleep stages.

Additionally, for Participant 1, S1 Sleep was significantly higher than S2 Sleep and REM Sleep. Within each Sleep and Wake category, there was considerable variation around the mean. This differed category-to-category but was considerably lower in SWS. This is likely due to the stable, recurrent, slow-wave brain activity associated with SWS, as well as the physical stability of deep sleep reducing noise in the recording.

These findings replicate those found in previous research, with LZc during conscious wake being significantly higher compared to during deep stages of sleep (Andrillon et al., 2016; Schartner et al., 2017a). This provides validity for our study and the subsequent results because it shows that even when using long-duration EEG and averaging over multiple days of data, the main effects established in previous research are still found. This was expected, given that previous studies often find the same general pattern of results overall, regardless of the finer details of study design or specific parameters. For example, often different forms of recording are used, including high-resolution EEG (Andrillon et al., 2016; Schartner et al., 2015), intracranial EEG (Schartner et al., 2017a) or magnetoencephalography (MEG) (Schartner et al., 2017b), yet the same trend of results is found.

Typically, research exploring the relationship between complexity measures, such as LZc, and consciousness has exclusively focused on group-level differences such as these. However, only the large effects between distinctly different global states of consciousness are usually reported. For example, in Schartner and colleagues' (2017a) sleep study, average LZc during an eyes-closed Wakeful Rest condition was presented as significantly higher than during SWS. However, this study neglected to report on the lighter sleep stages, S1 and S2 Sleep, citing that they were excluded to avoid "possible fluctuation of the level of consciousness due to subliminal processing" in these stages (Schartner et al., 2017a). Our results showed that S1 Sleep, S2 Sleep, and REM Sleep were all relatively similar in complexity level, sitting below Wake and above SWS. In P1, LZc during S1 was significantly higher than during S2 Sleep and

REM Sleep. This suggests a grouping of sleep stages in this middle ground, which LZc could not reliably differentiate.

Sleep stages were used in this study as a naturally occurring and predictable loss-of-consciousness (Hobson, 2008). This was ideal over the multi-day recording period. Different sleep stages are not necessarily different global states of consciousness, but they do consist of distinctly different brain activity identifiable in each stage (Andrillon et al., 2016). Schartner and colleagues (2017a) ignore their finding of Wakeful Rest and REM sleep being similar, by justifying them both as “states associated with consciousness” (Schartner et al., 2017a). Wake and REM sleep are commonly regarded as “activated brain states”, in comparison to non-REM deep sleep (Abásolo, Simons, da Silva, Tononi and Vyazovskiy, 2015), but REM and Wake are still distinctly different from each other and should be able to be differentiated by a measure of consciousness.

It is common in this field of research for measures to struggle to discriminate between subtler differences of global states of consciousness. For example, differentiating between MCS and UWS (Sitt et al., 2014) or transitional periods in and out of anaesthetic sedation (Wang et al., 2017). While LZc is consistently found to be significantly different between Wake and SWS, which may appear to be a success, the differences in brain activity between these two states can easily be differentiated by most measures, even by eye (Silber et al., 2007). If research continues only to report the distinct differences and ignore more subtle findings, there will consequentially always be a limited perspective and over-reaching conclusions about the successfulness of these measures.

LZc during conscious wakefulness is highly variable. To study variability of the LZc measure, we investigated LZc during conscious wakefulness. This period of data is an extended duration of time where the global state of consciousness is considered to be relatively stable (Hobson, 2008). We defined the waking range in our data as up to 14-hours per day, from 30-minutes after waking up in the morning to 30-minutes before falling asleep in the evening. This was relative to each participant's individual sleep schedule each day. This included recordings over multiple days; eight days total for Participant 1 and four days total for Participant 2. This unique dataset was achieved using long-duration EEG recording. This provided a comprehensive perspective of LZc to study the variability of this measure, unrelated to any changes in overall consciousness. If LZc only indicates consciousness, as commonly claimed, it should be stable during Wake, when an individual's global state of consciousness is also stable (Hobson, 2008).

There was a visible variation of LZc across the waking day. The spread of the raw LZc data was considerable yet consistent. Even when the data was smoothed across 5-minute moving averages, there was visible variation around the mean. In the raw data, there is a visible scattering of lower LZc values for both participants, reflected by a tail of low LZc scores and a negative skew. LZc even reached as low as .19 for Participant 1, and .27 for Participant 2, on a possible scale of 0-1. For reference, the mean LZc for Participant 1 during wake and SWS were .67 and .52 respectively. For Participant 2, these were .64 and .45. Wake and SWS exhibited the most considerable differences in mean LZc between the categories. Given this information, it raises concerns to see the LZc reach so low during Wake, even when smoothed to reduce noise. At the category level, the Coefficient of Variance of Wake was not largely different from that of the other categories, despite it being a very long duration and all-encompassing category compared to the other sleep stages. This suggests that the variability reported here is not exclusively a result of noise caused by movements during the day, but perhaps inherent noise in the measure, unspecific to the waking range. The cause of this variation, as well as the impacts of noise, will require further investigation.

As a result of the long-duration EEG recordings across time, the raw LZc data can be visualised in a unique and meaningful way. This revealed the visible scattering of low LZc values below the mean for both participants, providing an explanation for the overall negative skew. This tells us that there is a higher likelihood of low LZc values below the mean than above the mean. This negative skew of the data is highly important in the context of consciousness research, given that level of LZc is used to infer something about an individual's level of consciousness. Schartner and colleagues (2015; 2017a; 2017b) established that LZc reduced during anaesthetic sedation and sleep, compared to wakeful rest, but also increased during psychedelic states. They explicitly state that all together, "these results suggest an operationally useful one-dimensional scale for level of consciousness", which also "extends in both directions from the baseline state" (Schartner et al., 2017b). However, the results in the present study reached almost zero on some occasions, but this by no means implies that the person's consciousness reached almost zero. Previous research has attempted to claim an increase or decrease relative to baseline, but this baseline consciousness period is evidently unstable. This will be especially problematic for any future clinical applications of LZc. Additionally, this linear conceptualization of consciousness (increasing and decreasing along a single dimension), does not align with current theoretical understandings of the multidimensional nature of consciousness.

LZc of small sample durations is variable and unreliable. Given the variability of LZc identified during conscious wakefulness, it was necessary to then identify the extent this will impact its practical use as a measure. As part of the assumed stability of conscious wakefulness, it is assumed that the small sample durations used in research experiments will be representative of this range. Typically, only 5-15 minutes of EEG is recorded during each condition, and the average LZc of this sample is assumed to represent the LZc of this condition overall. In the present study, we calculated the mean LZc of small sample durations, ranging from 1- to 120-minutes in length. This created an extensive list of all possible mean LZc values which could be obtained during Wake for a sample of that duration. This was repeated for each sample duration size. Again, each participant's data was analysed separately. This was another comprehensive insight into LZc data made possible by the use of long-duration EEG. These sample means of varying durations were created to investigate what range of possible average LZc values could be obtained from the conscious wake data, and also to explore the effect of sample length.

Our results presented a highly variable range of mean LZc values obtained from these samples. It is not surprising that the variability of samples decreases as sample duration increases. However, it is concerning that, at the sample duration typically used in research (5-15 minutes), the range of possible mean LZc values obtained is highly variable. The shape of this curve suggests there may be some optimal duration that increases the reliability of the sample mean, whilst still being manageably short and still practical for empirical research. It is clear from these results, however, that average LZc of a recording shorter than 10-minutes is less reliable than even just slightly longer recordings of 20-minutes. Currently, there is no consensus or guideline for recording durations for EEG-based complexity analyses such as LZc. Further investigations using long-duration EEG may be used to establish such guidelines based on empirical evidence.

As an extension to this question, we tested whether average LZc significantly differed between one day and the next, within a participant. Here, each day of data included up to 14-hours of recording, a substantial period of time to average over. However, we found in both participants that there were significant differences in average LZc between two or more days. Considering that the main effects in this field of research are significant differences between consciousness conditions, it is concerning to find significant differences within a stable, conscious condition. This raises questions regarding the validity of the measure for its intended purpose. The extent of this sample reliability will require further investigation, in particular, to compare within-category variation to between-category variation. Regardless, there is sufficient evidence to demonstrate that as these samples are being used in research, we cannot be confident in the results they produce.

Samples during Wake are effectively used as a control condition in research. Any change in LZc as associated with a change in consciousness is in reference to a change from this baseline wakefulness condition. A period of wakeful rest, typically seated with eye-closed, is most commonly used (Schartner et al., 2015; Schartner et al., 2017a; Schartner et al., 2017b). However, it is acknowledged that this is not representative of the whole waking range, as different results are produced when a different baseline is used. When a cognitive task was used as the consciousness baseline, it was found that LZc during REM was significantly lower than Wake (Andrillon et al., 2016), unlike results identifying LZc during Wakeful West and REM to be comparable (Schartner et al., 2017a). It is known that the EEG recording differs when a person has their eyes closed compared to eyes open (Barry, Clarke, Johnstone, Magee & Rushby, 2007) and even LZc can significantly differ between these conditions (Farns et al., 2019).

An issue within this field is that results are being interpreted as increased or decreased relative to a conscious wakefulness baseline. However, until now, no one has measured this middle ground and the range of LZc produced. It has also not been established how reliable sample recordings from within this range may be, or what parameters might affect the sample means reported. Many factors can contribute to the variability of EEG complexity during Wake, and therefore, many factors should be considered in the choice of this experimental condition in research.

Limitations of Our Study

Participants. Due to the intensive long-duration recording process and the resulting large dataset to be analysed, we chose to focus our study on only two participants. Participant

1 was a PNES patient who underwent the recording for diagnostic purposes. Participant 2 was recruited as an additional healthy control dataset. While Participant 1 did experience a PNES event during this recording period, reported by the participant as a loss of time, automated seizure detection algorithms did not identify abnormal EEG activity during this period. This event was excluded from the dataset and it can confidently be asserted that this is not a confound to our results. Long-duration EEG is routinely recorded in these cases and may prove to be a unique pre-collected resource available for future research. However, it is still a patient population, and care must be taken when screening patients and data to ensure it is suitable for this research use. Demographic information, such as age, should also be considered in future. It is not yet understood how LZc output, on the scale of 0-1, is comparable across-subjects. Therefore, we chose to analyse the data from each participant individually. A sample size of two is small but allowed us to examine the finer details of LZc data across an extended period of time, within an individual. This suited the aims of this study. Future research should increase participant numbers in order to establish within- and between-subject effects.

Equipment. The portable EEG system used in this study is typically used for epilepsy monitoring and diagnostic recording. It is used in instances where patients are sent home for the duration of recording, as opposed to remaining in hospital confined to a bed and attached to a stationary EEG system (Chacko & McCullagh, 2014). The data for Participant 1 was recorded for this purpose, so a standard epilepsy monitoring set up was used. Participant 2 was therefore recorded with the same setup, resulting in comparable data. Future studies can adapt and optimise this equipment for applications in consciousness research.

This standard portable EEG setup involves electrodes that are fixed to the scalp using glue. This glue holds the electrodes quite firmly in place but is slightly irritating for participants and can be difficult to remove. However, this is necessary to maintain a stable connection. For the purpose of sleep scoring, it was a limitation that EOG and EMG channels were not included in this EEG set up. Without these channels, sleep scoring is less accurate. This is a known limitation that may result in an underestimation of total REM sleep and arousals during REM (AASM, 2007). There are several free channels on the recording system, so EOG and EMG, along with any other additional electrodes required, can easily be added in future.

The EEG signal was transmitted from the portable system, worn in a pack around the participant's waist, to a nearby server. However, its range is only 30 metres, resulting in some connection drop-out issues if the participant exceeded this range. P1 left their server at home, so there were periods of no-data during the day when they were out of range. P2 carried the server with them, contained in a portable briefcase set up, so EEG recordings were continuous for the whole duration of the study. This resulted in a different limitation because the video camera had to be set up each time they moved. Therefore, P2's video was intermittent despite the EEG being continuous. The video was not a crucial component to our analyses but did supplement the sleep scoring process. However, in future, video will likely serve as a necessary tool for coding behaviour and tracking variables of interest during the recording period. Both participants acquired video recording during sleep every night, which compensated for the missing EOG and EMG channels during the sleep scoring process.

Noise. With portable EEG recordings of this length, there are many factors that would contribute to noise in the data which are unrelated to neural activity or, unrelated to our variable of interest: level of consciousness. Noise refers to any data recorded that is not a meaningful signal recorded from neural activity. These artefacts in the signal are introduced by sources such as poor electrode connection, movement of electrodes, electricity mains interference, and physiological sources such as heart rate, muscle activity, eye movements and blinks (Reddy & Narava, 2013; Romo-Vázquez, Ranta, Louis-Dorr & Maquin, 2007). With EEG, the signal-to-noise ratio (SNR) is small. This means that there is a high level of noise in the data compared to the small level of meaningful signal (Reddy & Narava, 2013). This noise presents a challenged in all EEG analyses and interpretations, requiring strategies to prevent and remove noise in the data (Reddy & Narava, 2013). This is a common problem in all EEG but is typically minimised during recording with behavioural controls. For example, requiring participants to remain as still as possible during trials, and avoiding touching electrodes, moving their face, clenching their jaw, or blinking in the middle of trials (Reddy & Narava, 2013). In our study, behaviours were not monitored or controlled enough to avoid these sources of noise. However, the equipment used is more resilient to these artefacts than typical stationary EEG setups, due to the electrodes being held firmly in place with collodion glue (Gargiulo et al., 2010). The analysis that was employed in this experiment, whereby we calculated the LZc of 10-second segments, is also more resilient to these sources of noise than other EEG analyses. Typical EEG noise reduction and artefact rejection methods, such as trial-by-trial exclusions or filters (Romo-Vazquez et al., 2007), are not commonly used when conducting LZc analyses. Whilst there are methods to filter artefacts generated in long-duration EEG data (Boudet, Peyrodie, Forzy, Pinti, Toumi & Gallois, 2012), the effects of these filtering methods on LZc are unknown and were therefore not used in this study. Further investigation into the effect of noise on LZc, as well as the effect of noise-reduction filtering, should be completed in future.

Uncontrolled and unmonitored behaviour. An all-encompassing recording during conscious wakefulness was used in this study to represent the Wake condition. This decision was made because it was essential to have a comprehensive recording that could be used to represent this global state of consciousness. Conscious wakefulness is considered to be a relatively stable state of consciousness (Hobson, 2008). However, these long-duration recordings undoubtedly encompass a multifaceted mental, physiological, and behavioural experience. Given the purpose of this study, the procedure was purposefully kept simple. Participants were instructed to go about their daily lives, with no specific guidelines to their behaviour. This allowed a low-maintenance, comprehensive, and ecologically valid Wake condition. Video data was used to check that participants were in-range and not experiencing issues that might affect the equipment, although behaviour was not monitored or controlled. As part of the uncontrolled behaviour, the consumption of caffeine, nicotine, and alcohol was not restricted. This is a limitation as it may affect the EEG recording as well as complexity analysis (Chang, Song, Zhang, Shang, Ge & Wang, 2018). However, the intention was to keep the lives of the participants as consistent and normal as possible for this initial study. The main limitation of having unmonitored and undescribed behaviour is that it restricts the information and detail which can be drawn from our results. For the purpose of our aims, nothing else was required. However, it would be possible to introduce an experimental design to address this in future.

Investigative Framework

EEG-based complexity measures of consciousness. The results of the present study not only impact our understanding of LZc but all complexity-based measures of consciousness. LZc is an example of an EEG-based algorithmic complexity measure, but it one of many similar measures that have been implemented in this field of research. Conceptually, this genre of measures is based on complexity models of consciousness. These models propose that consciousness is necessarily supported by differentiation of neural activity (Tononi, 2008), functioning in some optimal range (Carhart-Harris, 2018). Therefore, complexity measures of spontaneous EEG are designed to quantify characteristics of differentiation in neural activity by calculating the number of dynamically distinct subsets of information within the EEG signal. Each measure utilises a different algorithmic approach to do so. For example, LZc applies a compression algorithm to count the number of unique subsequences of information in a data sequence, whilst Approximate Entropy (ApEn) utilises an entropy algorithm to quantify the regularity of data based on the predictability of a sequence (Ferenets, Lipping, Anier, Jäntti, Melto and Hovilehto, 2006; Gomez & Hornero, 2010). While computationally these measures calculate signal diversity via different means, conceptually they quantify similar characteristics in the signal (Ferenets et al., 2006).

The main effects identified are primarily the same across all measures; level of complexity is associated with the level of consciousness. While they differ slightly in their degree of accuracy, no one algorithm performs considerably better than the rest (Ferenets et al., 2006). Previous research has observed the performance of LZc as comparable to other complexity measures. Specifically, Schartner and colleagues (2015; 2017a; 2017b) tested LZc, Amplitude Coalition Entropy (ACE), and Synchrony Coalition Entropy (SCE) in their experiments, finding the same pattern of results for all three measures across all three experiments. There were subtle differences in the results of each measure but the main effects were the same. They also found a correlation between these three complexity measures, showing they are "similarly sensitive to certain signal features" (Schartner et al., 2017a). Other review papers find this same pattern (See Sitt et al. 2014), and experiments will often test multiple algorithms in the same study (Burns & Rjan, 2015). Schartner and colleagues (2017a) describe the different measures as capturing different "flavours" of complexity in the EEG recording. When multiple measures are compared, it is common for results to display the same effect from slightly different angles (Burns & Rjan, 2015; Ferenets et al., 2006). One measure alone is not as informative as multiple used in combination, suggesting a combined approach may provide more power and accuracy (Burns & Rjan, 2015; Schartner et al., 2017a; Sitt et al., 2014).

Research to establish the relationship between complexity measures and consciousness is all conducted in a similar way, often with multiple complexity algorithms tested alongside each other in the same experiments. Therefore, the same concerns and critiques discussed regarding LZc will apply to other EEG-based complexity measures of consciousness.

Designing a new algorithm will not bypass these issues. These are all empirical measures whereby the same empirical assumptions apply. The purpose of the present study was not to find evidence to establish or disprove LZc as a feasible measure, but to use LZc as an example to address issues which apply to all similar methods in this field. We hope to redirect the discussions currently held around the success of these measures and provide an empirical framework to address these issues directly.

Long-duration EEG. The success of the current study in identifying the variability of LZc has only been possible due to the empirical approach of using long-duration EEG recording. This method permitted consistent recording across extended periods of conscious wakefulness, as well as the whole sleep cycle, for multiple days. This is a novel application of portable long-duration EEG recording. This equipment is commonly utilised in epilepsy research and diagnosis (Ghougassian, D'Souza, Cook & O'Brien, 2004; Reuber & Elger, 2003), but applications in other research areas are relatively underexplored. With this study, we have shown the success of long-duration EEG in identifying the variability of LZc across time. This multi-day continuous recording is a rich and expansive dataset with the potential to provide unique and useful empirical insights in the field of consciousness research.

The main benefit of long-duration EEG is its ability to capture a comprehensive recording across a range of conditions of consciousness, instead of using small duration samples. The use of small duration samples is a limitation we identified in previous research, due to the shallow depth of insight and conclusions that could be made from the experimental results. Previous research can only claim significant differences in average complexity between distinct global states of consciousness conditions. In the present study, as with previous research, it was found that LZc during wake was significantly higher than during sleep. However, the use of long-duration EEG also facilitated the ability to expand, visualise, and describe the LZc output during conscious wakefulness. This exposed extensive variation;

information typically unavailable to conventional empirical studies. With this approach, not only are the conditions collected from a larger and more representative sample, but the data can be expanded and visualised across-time to reveal additional information.

A key outcome of the present study has been demonstrating the success of this methodological approach. This empirical approach allowed for direct observation of the variability of LZc, thereby directly addressing the assumption that LZc is stable during conscious wakefulness. This finding alone has a significant impact on our understanding of LZc. Regardless of the cause of this variability, its presence is a strong concern as it undermines core assumptions of this measure and its ability to indicate level of consciousness. Given that all EEG-based complexity measures of consciousness rely on this assumption, we propose they be tested in a similar way.

Complexity algorithms lend well to applications with long-duration EEG recordings. This is because the algorithms simply run over a series of data segments, of any given length, and provide a single output value for each segment. Additionally, given the ambiguous nature of consciousness, global states of consciousness are difficult to differentiate and categorise into discrete states (Bayne et al., 2016; Hobson, 2008). This approach provides greater flexibility to correlate with variables of interest, whilst also investigating and controlling for confounding variables and noise. This unique experimental approach can and should be used with all complexity-based measures to address and empirically test all underlying assumptions. Future research can improve the richness of information learned from the data by simply introducing more behavioural, physiological, and cognitive measures during the recording period. These can be correlated and compared with the complexity output, and additionally, be investigated across-time.

Reliability studies. In order to be established as empirical measures of consciousness in both research and practice, there are several factors these measures must empirically demonstrate. The purpose of this research is to learn about consciousness, as well as to ultimately employ these as tools in clinical practice. With any measure being developed towards this purpose, it is essential to thoroughly test and empirically establish its intended use. Here, we discuss a few examples of empirical evidence that a measure must demonstrate before claiming to indicate level of consciousness. Specifically, this is in reference to how a long-duration EEG framework has the unique ability to address these issues directly. These feasibility studies are necessary to show that measures are applicable in the contexts they are being marketed for, as well as ensuring sufficient evidence exists to support the conclusions being drawn about their capabilities.

Change in complexity should reflect a change in consciousness. A measure that is reliably indicative of level of consciousness should only change in relation to a corresponding change in level of consciousness. Conceptually, this is the aim and intention of all EEG-based complexity measures. However, the results of our study revealed variation in LZc during conscious wakefulness, a period of relatively stable consciousness, suggesting the measure varies considerably in response to other parameters. As part of establishing the relationship between consciousness and complexity, it is also necessary to thoroughly understand the influence that experimental parameters have on complexity output. This will influence the design of experiments, interpretation of results, and practical applications of the measures. A measure utilised to infer level of consciousness should, firstly, be indicative of changes in consciousness, but also, be largely unaffected by changes to other parameters. There are numerous possible sources of variation in EEG signal complexity and it is critical to identify and understand them before implanting such measures in practice. These may be related to different dimensions of cognition and consciousness, recording and analysis parameters, or simply caused by noise in the data.

Using long-duration EEG, it is possible to introduce additional measurements throughout a recording period, against which to correlate the changes in complexity. In our study, we purposefully used an all-encompassing recording of the waking range as our conscious condition. However, normal waking consciousness is comprised of many different aspects of cognition (Bayne et al., 2016; Hobson, 2008). Complexity is understood to significantly differ between eyes-open and eyes-closed EEG recordings (Farns et al., 2019) and the use of Wakeful Rest produces different results from a cognitive task, when used as a baseline condition (Andrillon et al., 2016; Schartner et al., 2017a). Therefore, within stable, conscious wakefulness, there are many potential parameters of consciousness within this global state that may relate to changes in complexity. The use of behavioural measures, sustained tasks, and self-reports of consciousness along various dimensions, may contribute to this understanding.

The effects of experimental and analysis parameters on complexity measures have been investigated in previous research. Experiments have, for example, tested altering segment duration, total number of electrodes, and location of electrodes. However, due to the limitations of typical experimental designs, the true impact of these parameters is not yet understood fully. Changes to parameters of the data analysis resulted in “no major changes” to the main observed effects and “broadly identical results” each time (Ferenets et al., 2006; Schartner et al., 2015; Schartner et al., 2017a; Schartner et al., 2017b). These results suggest that changes in complexity are robustly and reliably reflective of changes in consciousness. However, in all cases, the main effect retained is simply the significant difference between Wakeful Rest and some distinctly different conscious state. The details regarding how these changes affect the output of LZc are not reported on in any meaningful way. For example, the supplementary materials for Schartner and colleagues’ (2015) paper report that the effects found are robust to changes in duration of the segment size chosen. However, they only report that these measures are “robust across a large range of segment lengths”, with main effects lost for most participants with a segment size lower than 2-seconds (Schartner et al., 2015). The exact nature of how this changes the output of LZc, ACE, and SCE is unreported. A deeper understanding of these changes is needed before it can be concluded that they do not impact the overarching experimental effect.

The long-duration EEG framework applied in our experiment is an ideal approach to address these issues. The richness of understanding that is available within the data can be easily improved upon by implementing additional experimental parameters. Some examples include behavioural data collection, self-reports of levels of consciousness, cognitive task conditions, controls for confounding variables that might introduce noise, as well as periods of sustained and stable activities. Thorough investigations into the influence of altering analysis parameters should also be investigated, to rule out alternative explanations for the variation observed. Regardless of what causes the variability, and whether we can learn to predict it, the fact that LZc was found to vary without a corresponding change in global state of consciousness raises concern.

Complexity should be unaffected by noise. A large source of noise in EEG recordings is physical and can easily be induced by participant movements and behaviours, effecting the EEG signal recorded (Reddy & Narava, 2013; Romo-Vazquez et al., 2007). Given that different global states of consciousness are typically defined by different physical behaviours, as well as cognitive behaviours, it is crucial to understand how these differences may be contributing to complexity. This is necessary to ensure that changes in complexity are selectively sensitive to changes in consciousness. Behaviourally induced activity can even be significantly greater than underlying brain-related activity (Gwin, Gramann, Makeig & Ferris, 2010). This will be a particular issue with portable EEG, even in controlled environments or with monitored behaviour.

Muscle activity typically results in high-frequency information in the EEG signal, overlapping with high-frequency neural activity (Muthukumaraswamy, 2013). For depth of anaesthesia research, a low-pass filter of around 20Hz is often used, which eliminates most of the muscle activity in higher frequencies (Rautee, Sampson, Sarkela, Melto, Hovilehto & van Gils, 2005). In the present study, we did not filter to remove these artefacts. However, there are several ways being developed to address them, which may be used in future. Boudet and colleagues (2012) developed a means of filtering artefacts from long-duration EEG data. Their method uses Adaptive Filtering by Optimal Projection (AFOP), which is an automated method to address artefacts caused by eye movements, blinks, muscle activity, or movements in the electrodes (Boudet et al., 2012). However, this is being optimised for long-duration epilepsy monitoring and it is not clear how these artefacts and their removal impact complexity measures. Gwin and colleagues (2010) looked at a two-step approach to remove artefacts from data recorded while subjects were walking and running on a treadmill. They identified and removed artefacts that were phase-locked to stride and gait of the walk or run. However, only a small portion of this artefact could be removed, given the complicated impact on the EEG recording (Gwin et al., 2010). Artefact studies such as these should be completed with complexity analyses, to improve the SNR in the data.

A measure that is primarily impacted by noise cannot be solely indicative of level of consciousness. The higher complexity of Wake compared to DOC states should not only be reflective of the difference in movement between the two population groups, as this does not disclose anything useful regarding the internal experience of that person. On the one hand, it is important to understand and predict the sources of noise in your data, but on the other hand, a measure of consciousness such as these should be selectively sensitive to changes in neural activity associated with changes in consciousness.

Complexity must be proven across all forms of altered consciousness. Testing has been completed across a range of alterations of consciousness. For example, sedation monitoring with different anaesthetic drugs (e.g. Bruhn, Röpcke & Hoeft, 2000; Schartner et al., 2015), differentiating between DOC (e.g. Sitt et al., 2014), stages of sleep and dreaming (e.g. Burioka et al., 2005; Fell, Röschke, Mann & Schäffner, 1996; Schartner et al., 2017a), psychedelic drugs (e.g. Schartner et al., 2017b), and visual hallucination experiences (Schwartzman, Schartner, Ador, Simonelli, Change & Seth 2019). The neural mechanisms altering these different cases can be very different. For example, sleep is a naturally occurring loss of consciousness, with different sleep stages that are identifiable via distinct spectral profiles (Silber et al., 2007). Anaesthetic drugs work via specific neurochemical mechanisms that are dependent on the type of drug, typically by suppressing GABAergic pathways in the brain, which results in decreased cortical activity (Voss and Sleigh, 2007). Psychedelic drugs are known to alter subjective conscious experience resulting in altered content of consciousness (Vollenweider & Kometer, 2010). DOC occur as a result of severe neural injury across a wide range of regions in the brain (Owen, 2008).

It is not yet understood how the neural changes associated with these distinctly different mechanisms can alter the diversity of signal recorded and computed via complexity measures. Ferenets and colleagues (2006) suggest that perhaps the use of these measures should be developed more specifically to each case of altered consciousness, instead of trying to develop an overarching measure of consciousness more universally. They also suggest a more detailed analysis of how these measures behave in each context is needed. While complexity measures intend to locate commonality between all of these cases, this must be empirically tested and established before simply applying across the board. These measures must be empirically established in each specific case, not just assumed to generalise.

Must provide an interpretation of the metric to be used. An overarching aim in this field is to use these measures in practice to infer information about level of consciousness in an individual. Ferenets and colleagues (2006) state that the most significant barrier to using these measures in clinical practise regards the difficulty of interpreting the results. A significant issue with EEG-based complexity measures of consciousness is that the output of these measures is not on an objective scale. Experiments typically apply an objective measure of effect size to compare between participants, without using or even discussing the numerical scales which these measures output. Some studies have shown complexity to successfully differentiate between UWS and MCS, for example, but there is no classification system like there is with the PCI metric (Sitt et al., 2014). In contrast, it is claimed that PCI is an objective scale, with the ability to provide a clear distinction between conscious and unconsciousness in individual, drawing a line between them (Casarotto et al., 2016). Developing classifiers is particularly difficult because the behavioural assessments that are relied upon to train the classifiers are themselves unreliable**.**

The importance of our approach in the present study was to analyse the LZc within an individual across an extended period of time. By doing so, it was possible to focus on how LZc output varies along the 0-1 scale and identify important features. For example, we observed LZc reaching almost zero at multiple times throughout the day. We were also able to identify a similar pattern of results across both participants but shifted up/down on the LZc scale. The only way to utilise this as a metric in clinical practice and to inform knowledge of level of consciousness, would be to have a method of interpreting the measure that is comparable across people. Otherwise, it would be necessary to establish before and after comparison conditions to calculate the change in complexity within an individual. This approach may be possible for applications such as tracking depth of anaesthesia, for example, but is not useful for diagnosing DOC patients.

Studies to establish a reliable measure, even within a person, would need to confirm that complexity is stable within a similarly stable conscious state. A complexity measure must show it is reliably the same when measuring from one day to the next, from the same person in the same state. The results of our study showed that LZc was not stable day-to-day; evidence that there is no intra-subject reliability. While the interpretation of the results must be established, the practical idea of attaching an EEG cap and leaving it to record for a few days has extensive clinical appeal. These measures have considerable potential to be useful in practical applications, but this further reinforces the importance of establishing groundwork empirical evidence to support these uses.

The Concept of Measuring Consciousness

'Levels' of consciousness. Global states of consciousness are the overall conditions of an integrated conscious experience. They are typically defined by their general cognitive, behavioural, and physiological experiences, rather than by specific contents or individual cognitive elements (Bayne et al., 2016; Hobson, 2008). Conscious Wakefulness is considered to be a relatively stable state of consciousness and a baseline against which to judge all alterations of consciousness (Hobson, 2008). Several different things can cause deviations from this baseline. For example drugs, psychosis, sleep, or meditation (Hobson, 2008). While not always discrete 'states' necessarily, states of consciousness are considered to be periods of relative stability in these dimensions as well as relative stability of experience (Tononi, 2008).

In an attempt to capture all global states of consciousness, including different altered conscious states, researchers have begun to consider these states as different 'levels' of consciousness. For example, the set of studies by Schartner and colleagues (2015; 2017a; 2017b) compare results from sedation, sleep, and psychedelics. Results presenting reduced complexity compared to a conscious baseline (Schartner et al., 2015; Schartner et al., 2017a), and increased complexity comparative to this baseline (Schartner et al., 2017b), are interpreted as changes in levels of consciousness along a single conceptual one-dimensional scale. A significant motivation of the current study was to address the lack of comprehensive studies or recording surrounding the baseline of conscious wake, because its upper and lower limits are unknown. Results of the present study revealed that there is a high level of variability during conscious wake, thus impacting the interpretation of previous research and the apparent increase or decrease relative to this apparent baseline.

Schartner and colleagues (2017b) present increased complexity during psychedelic states, compared to a placebo control group, and were the first to display an increase in complexity compared to normal waking consciousness. They interpret this to suggest that the state of consciousness induced by psychedelic drugs is an “elevated level of consciousness” (Schartner et al., 2017b). However, they do warn that describing this as a “higher” state consciousness is dubious and that these interpretations will need to be adequately established (Schartner et al., 2017b). Bayne and Carter (2018) argue that while specific aspects of cognition and perception are enhanced, many other cognitive capacities are impaired, including many associated with consciousness. Therefore, while an altered state of consciousness, they say it is "inappropriate to regard them as 'higher' states of consciousness". Bayne and Carter (2018) argue that this evidence supports a multidimensional understanding of consciousness, as opposed to a levels-based conception. A study by Schwartzman and colleagues (2019) investigated the signal diversity of a non-pharmacologically altered state of consciousness. They induced visual hallucinations using stroboscopic stimulation to alter subjective experience. This identified significantly increased LZc for both stimulation intensity condition and these results were interpreted as showing LZc to be indicative of diversity of subjective experience (Schwartzman et al., 2019).

Describing global states of consciousness as ‘levels’ of consciousness that can be roughly ordered along a single dimension is seemingly appealing, especially for clinical contexts. Nonetheless, it is problematic at a conceptual level (Bayne et al., 2016). This idea has become more than just an informal ordering and Bayne and colleagues (2016) argue that the “central theoretical construct” of levels still must be discussed. This is a theoretical issue, to begin with, but important practical issues will emerge when these ideas allow labelling of one person as ‘more’ or ‘less’ conscious based on these results, for example. There are also numerous aspects of consciousness that shift and change within a state, which may not be captured by this conceptualisation of levels. Bayne and colleagues (2016) propose a shift from ‘levels’ to multidimensional states, where states are regions of stability within this multidimensional state space. Currently, our conceptual understanding of how these global states of consciousness are understood in relation to each other is underdeveloped.

Relationship to the concept of complexity. This field of research is based on the theoretical foundation that complexity of neural activity underlies conscious experience. This forms an assumption that a more conscious brain would have less organised activity and vice versa. The ordering of conscious states may be convenient, but the interpretation of the complexity metrics does not necessarily align with the theoretical concept, at a practical level. There is an existing discussion regarding ‘higher’ or ‘lower’ levels of consciousness, but what is often ignored is that very high and very low complexity are both equally meaningless. With LZc, for example, 0 is completely compressible, and 1 is completely random. Both are not conducive to the type of activity required to support a conscious experience. It is more accurate to consider consciousness existing in some optimal middle ground (Carhart-Harris, 2018), instead of a linear relationship between increased consciousness and increased complexity. Therefore, it seems clear that the increased complexity observed in psychedelic states (Schartner et al., 2017b) and during stroboscopically-induced visual hallucinations

(Schwartzman et al., 2019), should not be assumed to be ‘higher’ levels of consciousness. The increased complexity in these conditions is more likely to be reflective of a more diverse, yet not necessarily more conscious experience (Bayne & Cartner, 2018).

If we acknowledge and understand how these complexity measures work, we know that anything that increases organised activity will reduce complexity in the signal. For example, the tail of low LZc scores observed during Wake in the present study could likely be caused by physical sources of noise, not fluctuations in consciousness. In the case of this multi-channel LZc calculation, anything that increases common signal across multiple electrodes will also reduce complexity (Schartner et al., 2015). Therefore, repetitive muscle activity during chewing or talking will induce a common signal across all electrodes. We speculate this is the cause of the low LZc values observed during Wake, and consider it an example of how a source of noise can impact complexity, based on a knowledge of how this LZc compression algorithm works.

A measure may be intending to indicate level of consciousness, but if it does so via measuring something else, then it is not quantifying anything directly relevant to consciousness, per se. While an EEG-based measure for monitoring depth of sedation may actually benefit from sensitivity to muscle activity, as this would increase with patient arousal (Ferenets et al., 2006), these are not direct measures of consciousness and do not align with what complexity measures are intended for. Complexity measures of consciousness are marketed as superior to other methods due to the premise that they measure a characteristic of neural activity directly relevant to conscious experience. However, there is still much to discuss in regards to whether EEG can accurately capture neural diversity in the way it is intended. It is important that the EEG recording is actually detecting characteristics of neural activity that are theoretically relevant to consciousness, and that the complexity algorithms are sensitive to this diversity, also. The aim is to discover a measure that is selectively sensitive to consciousness. Considering this, it is clear why LZc is problematic as a measure solely indicative of consciousness, due to its propensity to vary without any change in conscious condition.

Clinical Implications and Significance of Findings

With the application of neuroimaging to DOC research, it has become essential to discuss the ethical use of these tools in practice (Murphy et al., 2008). In particular, a common overarching aim of this research field is to improve DOC diagnosis practices. Current methods of diagnosis are limited to unreliable behavioural and physiological assessments (Giacino, Kalmar & Whyte, 2004; Schnakers et al., 2009). This is a vulnerable population and a complex ethical situation, where neuroimaging methods have the potential to help provide evidence towards more appropriate care. However, this also comes with a lot of responsibility, as there are significant consequences if this evidence is misleading and leads to inappropriate care. If it is claimed that a measure is able to indicate or differentiate between global states of consciousness, it may in future be used to inform decisions regarding the consciousness of patient groups.

Decisions regarding ongoing care for DOC patients are made with considerations for quality of life, primarily based on the patient's level of awareness and cognitive abilities. Current methods, including behavioural assessments and physiological signs, unfortunately, provide very limited insight into awareness and cognition (Giancino et al., 2004; Owen, 2008; Seel et al., 2010). It has been shown that a portion of patients diagnosed with UWS actually retain more awareness than previously believed (Owen et al., 2006). Evidence like this is critical because in all severe DOC cases, there are clinical decisions made regarding whether to continue life-supporting care (Jennet, 2005). The costs of maintaining care and life support measures are weighed against the benefits, and care may be withdrawn out of respect for the patient and their family's suffering (Jennet, 2005; Murphy et al., 2008). The long-term outcomes are always decided within the first year after injury and may result in the diagnosis of a permanent UWS state. It is often thought that there is no benefit to prolonging the life of someone in a permanent UWS state (Jennet, 2005). Therefore, there is significant weight to these decisions. Science currently falls short of understanding DOC enough to make accurate calls (Hirsch, 2005), yet it is up to doctors to make these decisions regularly, despite incomplete understanding (Jennet, 2005).

EEG-based complexity measures may have the potential to provide useful information regarding the overall level of awareness in patients. Unlike PCI, which aims to categorise patients into conscious and unconscious groups (Casarotto et al., 2016), spontaneous EEG complexity measures may provide a more useful description of consciousness across-time. Fischer & Truog (2017) argue that with DOC, it is concerning to differentiate MCS patients as 'conscious' and UWS patients as 'unconscious'. This distinction is not reasonable. At a conceptual level, conscious and unconscious are not binary states, and at a practical level, it is difficult to truly know a person's internal consciousness experience (especially with limited measures to provide insight). Fischer and Troug (2017) encourage the focus to be, instead, on the quality and nature of a patient's experience (Fischer & Truog, 2017). Investigations utilising long-wear EEG can provide insights into dynamics within a person across time. This is useful diagnostic information because fluctuations in signs of awareness are criteria for differentiating MCS from UWS, with MCS patients exhibiting fluctuations of increased awareness across the day (Schiff et al., 2014). To capture these fluctuations with static measurement methods would require repeated measurements throughout the day (Noirhomme, Brecheisen, Lesenfants, Antonopoulos & Laureys, 2015). This is not cost or time effective and relies on chance to be measuring at the right time to capture a fluctuation in consciousness.

As has been demonstrated in this study, complexity analyses can successfully be run over continuous EEG data recorded over long-durations of time. This is appealing, as it can be attached to the patient and left recording for multiple hours with no active monitoring needed from physicians. This is a suitable tool for this application. However, it must still be empirically tested for this use. The results of our study showed variability across a day within conscious participants. Further research would be required to establish whether this variability exists in all patient groups and conscious conditions, or if, potentially, the fluctuations in complexity are insightful in some way. The use of complexity measures across-time could also be used to map sleep. Sleep cycles, including circadian rhythms, are an important diagnostic tool for DOC, typically marking an emergence from a coma to UWS (Bekinschtein et al., 2009; Blume et al., 2015). However, these sleep cycles are difficult to identify in EEG signal using typical frequency-based sleep scoring methods (Blume et al., 2015).

Currently, there are no worldwide standards of care for DOC, partly due to there being no established treatments, or measures of long-term changes to condition (Murphy et al., 2008; Giacino & Whyte, 2004). Developing measures that indicate level of consciousness across-time are vital to being able to identify if there are changes in a patient’s condition. This sort of assessment is necessary to be able to study the effectiveness of treatments because improvements would only occur over a long period of time (Laureys et al., 2010).

Given the significant ethical weight of using tools to improve diagnostic accuracy in DOC patients, there are also significant consequences to using tools we do not yet fully understand. There is much appeal to having neuroimaging evidence supporting a diagnosis and care plan. Results from our study showed there is a negative skew in LZc, resulting from a scattering of low LZc scores throughout the day, reaching close to 0. This is vital information to know because it may result in estimations of lower complexity. The consequences of underrating patients’ awareness levels with inaccurate diagnoses could, at the extreme, contribute to premature withdraws of life-supporting care (Jennet, 2005). A measure that does not work as it is claimed to work is worse than no measure at all. This is why, in this study, we have strongly emphasised the need to thoroughly test complexity measures that have been claimed to be able to indicate level of consciousness.

As an extension to these concerns, there are other cases where EEG complexity, as an indicator of level of consciousness, may be used inappropriately and prematurely. For example, it was recently shown that EEG complexity (LZc, ACE, SCE) can increase with psychedelic states, claimed to reflect a higher level of consciousness and increased diversity of subjective experience (Schartner et al., 2017b; Schwartzman et al., 2019). Scott and Carhart-Harris (2019), inspired by these results, published a paper discussing the idea of giving psychedelic drugs to DOC patients to increase their level neural complexity and, therefore, increase their level of consciousness. This paper was just a published discussion regarding the ethical considerations of completing a study like this, however, it is a very problematic leap in logic that is not supported by current evidence. This is significant cause for concern and highlights how close these experiments may be to happening.

It is important to recognise both the value and limitations of neuroimaging methods as indicators of levels of consciousness. Currently, evidence does not support their use in clinical practice (Jox et al., 2012). While clinical applications are a key motivation for the development of these measures, there is substantial research that must be completed before they can be implemented in practice. It is also essential to maintain a dialogue regarding the ethical considerations of applying measures in this field, in particular, to the diagnostic decisions of DOC patients (Hirsch, 2005). In the present study, we have highlighted the importance of empirically testing the assumptions and claims made in regards to complexity measures of consciousness, regardless of their intended applications.

Conclusion

Complexity measures of consciousness stem from the concept that consciousness is supported by integrated and differentiated neural activity (Tononi, 2008), whereby there is some optimal balance of neural activity that is diverse yet organised enough to support conscious experience (Carhart-Harris, 2018). Algorithms from Information Theory have been utilised to quantify these characteristics in spontaneous EEG signal. LZc is one example of such measures, which applies a compression algorithm to quantify the diversity in a given segment of multi-channel EEG data (Schartner et al., 2015). The use of spontaneous EEG provides many practical benefits compared to other neuroimaging methods, such as fMRI or PCI. It is low-cost, can be implemented bed-side, and provides a unique perspective of recording the dynamics of neural activity across time.

Across all EEG-based complexity measures, including LZc, findings suggest that the level of complexity is reflective of the level of consciousness. However, there is a significant lack of knowledge about how these measures may vary, either with or without a corresponding change in consciousness. Previous study designs compare a baseline healthy wakeful consciousness condition with an altered consciousness condition; therefore, inferring that the difference in complexity observed between conditions is reflective of a difference in global state of consciousness. However, this relies upon the assumption that complexity during this conscious baseline condition is relatively stable, and that the small sample recordings (e.g. 5-10 minutes) from this condition are representative of complexity during conscious wakefulness.

To directly address this assumption, we used the unique methodological approach of portable, long-duration EEG recording. Our dataset consisted of multiple days and nights of continuous EEG recorded from two participants (8 days total for Participant 1, 4 days total for Participant 2) and with individual analysis of each participants' data. No behavioural manipulations or controls were enforced during this recording period; participants were sent home to carry out their daily lives as normal. LZc was then computed across the whole duration of continuous recordings, providing a single complexity value for every 10-second segment.

Although we replicated previous findings of LZc as significantly lower during Sleep than Wake, results from this study showed that LZc varied considerably within Wake, a period of relatively stable consciousness. This variability impacted the mean LZc identified using samples of smaller duration (as is typically used in research). Although this variability reduced with a longer sample duration, LZc from one whole day could also significantly differ from another day within the same subject. Regardless of the source of the variation, this is evidence to suggest that LZc can vary considerably whilst consciousness remains relatively stable.

Given the clinical imperative to develop a brain-based measure of consciousness that is both objective and reliable, this variability is of significant concern. The variation suggests that LZc may not be indicative of level of consciousness in an individual, as previously claimed. Using a tool for measuring consciousness that we do not fully understand has potentially significant clinical and ethical implications. The issues raised and addressed in this study are not unique to LZc but will apply to all similar EEG-based complexity measures of consciousness. They will not be solved by merely improving algorithm design or methods of analysis. Instead, all complexity measures should undergo reliability testing as both a proof of concept and a proof of practice. The motivations to develop a measure that is indicative of consciousness are important and it would be of significant ethical concern to use a measure that has not been empirically established for its intended applications.

The use of long-duration EEG in this study successfully facilitated the identification of variability in LZc in a way that standard experimental study designs could not, demonstrating that this experimental framework is successful for this purpose. We have shown that this method can provide an information-rich dataset which is uniquely capable of exposing variability in the complexity of spontaneous EEG signal. Additionally, we have discussed the potential of using this approach to address other assumptions and investigate the nature of complexity measures across time. We hope to redirect discussions in this field to these critical issues and promote the completion of similar reliability studies using this framework. These studies can and should be conducted with all complexity measures of consciousness, to ensure a reliable measure of consciousness can be developed for future clinical and research use.

References

Abásolo, D., Simons, S., Morgado da Silva, R., Tononi, G., & Vyazovskiy, V. (2015). Lempel-Ziv complexity of cortical activity during sleep and waking in rats. *Journal Of Neurophysiology*, *113*(7), 2742-2752. doi: 10.1152/jn.00575.2014

Andrillon, T., Poulsen, A., Hansen, L., Léger, D., & Kouider, S. (2016). Neural Markers of Responsiveness to the Environment in Human Sleep. *The Journal Of Neuroscience*, *36*(24), 6583-6596. doi: 10.1523/jneurosci.0902-16.2016

Bar, M., Tootell, R., Schacter, D., Greve, D., Fischl, B., & Mendola, J. et al. (2001). Cortical Mechanisms Specific to Explicit Visual Object Recognition. *Neuron*, *29*(2), 529-535. doi: 10.1016/s0896-6273(01)00224-0

Barry, R., Clarke, A., Johnstone, S., Magee, C., & Rushby, J. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, *118*(12), 2765-2773. doi: 10.1016/j.clinph.2007.07.028

Bayne, T. (2011). The Presence of Consciousness in Absence Seizures. *Behavioural Neurology*, *24*(1), 47-53. doi: 10.1155/2011/128567

Bayne, T., & Carter, O. (2018). Dimensions of consciousness and the psychedelic state. *Neuroscience Of Consciousness*, *2018*(1). doi: 10.1093/nc/niy008

Bayne, T., Hohwy, J., & Owen, A. (2016). Are There Levels of Consciousness? *Trends In Cognitive Sciences*, *20*(6), 405-413. doi: 10.1016/j.tics.2016.03.009

Bender, A., Jox, R., Grill, E., Straube, A., & Lulé, D. (2015). Persistent Vegetative State and Minimally Conscious State. *Deutsches Aerzteblatt Online*, *112*(14), 235-242. doi: 10.3238/arztebl.2015.0235

Blume, C., del Giudice, R., Wislowska, M., Lechinger, J., & Schabus, M. (2015). Across the consciousness continuum - from unresponsive wakefulness to sleep. *Frontiers In Human Neuroscience*, *9*(105). doi: 10.3389/fnhum.2015.00105

Boudet, S., Peyrodie, L., Forzy, G., Pinti, A., Toumi, H., & Gallois, P. (2012). Improvements of Adaptive Filtering by Optimal Projection to filter different artifact types on long duration EEG recordings. *Computer Methods And Programs In Biomedicine*, *108*(1), 234-249. doi: 10.1016/j.cmpb.2012.04.005

Bruhn, J., Röpcke, H., & Hoeft, A. (2000). Approximate Entropy as an Electroencephalographic Measure of Anesthetic Drug Effect during Desflurane Anesthesia. *Anesthesiology*, *92*(3), 715-726. doi: 10.1097/00000542-200003000-00016

Bruno, M., Vanhaudenhuyse, A., Thibaut, A., Moonen, G., & Laureys, S. (2011). From unresponsive wakefulness to minimally conscious PLUS and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness. *Journal Of Neurology*, *258*(7), 1373-1384. doi: 10.1007/s00415-011-6114-x

Burioka, N., Miyata, M., Cornélissen, G., Halberg, F., Takeshima, T., & Kaplan, D. et al. (2005). Approximate Entropy in the Electroencephalogram during Wake and Sleep. *Clinical EEG And Neuroscience*, *36*(1), 21-24. doi: 10.1177/155005940503600106

Burns, T., & Rajan, R. (2015). Combining complexity measures of EEG data: multiplying measures reveal previously hidden information. *F1000research*, *4*(137), 137. doi: 10.12688/f1000research.6590.1

Carhart-Harris, R. (2018). The entropic brain - revisited. *Neuropharmacology*, *142*, 167-178. doi: 10.1016/j.neuropharm.2018.03.010

Casali, A., Gosseries, O., Rosanova, M., Boly, M., Sarasso, S., & Casali, K. et al. (2013). A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*, *5*(198). doi: 10.1126/scitranslmed.3006294

Casarotto, S., Comanducci, A., Rosanova, M., Sarasso, S., Fecchio, M., & Napolitani, M. et al. (2016). Stratification of unresponsive patients by an independently validated index of brain complexity. *Annals Of Neurology*, *80*(5), 718-729. doi: 10.1002/ana.24779

Chacko, S., & McCullagh, P. (2014). Home Based Mobile Solution for Video Ambulatory EEG Monitoring. *2014 International Conference On Intelligent Environments*. doi: 10.1109/ie.2014.23

Chalmers, D. (1995). Facing Up to the Problem of Consciousness. *Journal Of Conscious Studies*, *2*(3), 200-219. doi: 10.1093/acprof

Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T., Fernández-Espejo, D., & Pickard, J. et al. (2011). Bedside detection of awareness in the vegetative state: a cohort study. *The Lancet*, *378*(9809), 2088-2094. doi: 10.1016/s0140-6736(11)61224-5

Dehaene, S., & Changeux, J. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200-227. doi: 10.1016/j.neuron.2011.03.018

Dehaene, S., Changeux, J., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends In Cognitive Sciences*, *10*(5), 204-211. doi: 10.1016/j.tics.2006.03.007

Farnes, N., Juel, B., Nilsen, A., Romundstad, L., & Storm, J. (2019). Increased signal diversity/complexity of spontaneous EEG, but not evoked EEG responses, in ketamine-induced psychedelic state in humans. doi: 10.1101/508697

Fell, J., Röschke, J., Mann, K., & Schäffner, C. (1996). Discrimination of sleep stages: a comparison between spectral and nonlinear EEG measures. *Electroencephalography And Clinical Neurophysiology*, *98*(5), 401-410. doi: 10.1016/0013-4694(96)95636-9

Ferenets, R., Lipping, T., Anier, A., Jantti, V., Melto, S., & Hovilehto, S. (2006). Comparison of Entropy and Complexity Measures for the Assessment of Depth of Sedation. *IEEE Transactions On Biomedical Engineering*, *53*(6), 1067-1077. doi: 10.1109/tbme.2006.873543

Ferenets, R., Vanluchene, A., Lipping, T., Heyse, B., & Struys, M. (2007). Behavior of Entropy/Complexity Measures of the Electroencephalogram during Propofol-induced Sedation. *Anesthesiology*, *106*(4), 696-706. doi: 10.1097/01.anes.0000264790.07231.2d

Fingelkurts, A., Fingelkurts, A., Bagnato, S., Boccagni, C., & Galardi, G. (2014). Do we need a theory-based assessment of consciousness in the field of disorders of consciousness? *Frontiers In Human Neuroscience*, *8*. doi: 10.3389/fnhum.2014.00402

Fins, J., Illes, J., Bernat, J., Hirsch, J., Laureys, S., & Murphy, E. (2008). Neuroimaging and Disorders of Consciousness: Envisioning an Ethical Research Agenda. *The American Journal Of Bioethics*, *8*(9), 3-12. doi: 10.1080/15265160802318113

Fischer, D., & Truog, R. (2017). The Problems With Fixating on Consciousness in Disorders of Consciousness. *AJOB Neuroscience*, *8*(3), 135-140. doi: 10.1080/21507740.2017.1366577

Gargiulo, G., Calvo, R., Bifulco, P., Cesarelli, M., Jin, C., Mohamed, A., & van Schaik, A. (2010). A new EEG recording system for passive dry electrodes. *Clinical Neurophysiology*, *121*(5), 686-693. doi: 10.1016/j.clinph.2009.12.025

Ghahramani, Z. (2006). Information Theory. *Encyclopedia Of Cognitive Science*. doi: 10.1002/0470018860.s00643

Ghougassian, D., D'Souza, W., Cook, M., & O'Brien, T. (2004). Evaluating the Utility of Inpatient Video-EEG Monitoring. *Epilepsia*, *45*(8), 928-932. doi: 10.1111/j.0013-9580.2004.51003.x

Giacino, J., Fins, J., Laureys, S., & Schiff, N. (2014). Disorders of consciousness after acquired brain injury: the state of the science. *Nature Reviews Neurology*, *10*(2), 99-114. doi: 10.1038/nrneurol.2013.279

Giacino, J., Kalmar, K., & Whyte, J. (2004). The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Archives Of Physical Medicine And Rehabilitation*, *85*(12), 2020-2029. doi: 10.1016/j.apmr.2004.02.033

Goldfine, A., Victor, J., Conte, M., Bardin, J., & Schiff, N. (2011). Determination of awareness in patients with severe brain injury using EEG power spectral analysis. *Clinical Neurophysiology*, *122*(11), 2157-2168. doi: 10.1016/j.clinph.2011.03.022

Gómez, C., & Hornero, R. (2010). Entropy and Complexity Analyses in Alzheimer’s Disease: An MEG Study. *The Open Biomedical Engineering Journal*, *4*(1), 223-235. doi: 10.2174/1874120701004010223

Gosseries, O., Thibaut, A., Boly, M., Rosanova, M., Massimini, M., & Laureys, S. (2014). Assessing consciousness in coma and related states using transcranial magnetic stimulation combined with electroencephalography. *Annales Françaises D'anesthésie Et De Réanimation*, *33*(2), 65-71. doi: 10.1016/j.annfar.2013.11.002

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*(8), 837-843. doi: 10.1038/77754

GuruvaReddy, A., & Narava, S. (2013). Artifact Removal from EEG Signals. *International Journal Of Computer Applications*, *77*(13), 17-19. doi: 10.5120/13543-1175

Gwin, J., Gramann, K., Makeig, S., & Ferris, D. (2010). Removal of Movement Artifact From High-Density EEG Recorded During Walking and Running. *Journal Of Neurophysiology*, *103*(6), 3526-3534. doi: 10.1152/jn.00105.2010

Hirsch, J. (2005). Raising consciousness. *Journal Of Clinical Investigation*, *115*(5), 1102-1102. doi: 10.1172/jci25320

Hobson, J. (2008). Normal and Abnormal States of Consciousness. In M. Velmans & S. Schneider, *The Blackwell Companion to Consciousness*. Wiley Blackwell.

Jennett, B. (2005). Thirty years of the vegetative state: clinical, ethical and legal problems. *Progress In Brain Research*, 537-543. doi: 10.1016/s0079-6123(05)50037-2

Jox, R., Bernat, J., Laureys, S., & Racine, E. (2012). Disorders of consciousness: responding to requests for novel diagnostic and therapeutic interventions. *The Lancet Neurology*, *11*(8), 732-738. doi: 10.1016/s1474-4422(12)70154-0

Kissin, I. (2000). Depth of Anesthesia and Bispectral Index Monitoring. *Anesthesia & Analgesia*, *90*(5), 1114-1117. doi: 10.1097/00000539-200005000-00021

Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: progress and problems. *Nature Reviews Neuroscience*, *17*(5), 307-321. doi: 10.1038/nrn.2016.22

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends In Cognitive Sciences*, *14*(7), 301-307. doi: 10.1016/j.tics.2010.04.006

Lamme, V., & Roelfsema, P. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends In Neurosciences*, *23*(11), 571-579. doi: 10.1016/s0166-2236(00)01657-x

Lempel, A., & Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions On Information Theory*, *22*(1), 75-81. doi: 10.1109/tit.1976.1055501

Massimini, M., Ferrarelli, F., Esser, S., Riedner, B., Huber, R., & Murphy, M. et al. (2007). Triggering sleep slow waves by transcranial magnetic stimulation. *Proceedings Of The National Academy Of Sciences*, *104*(20), 8496-8501. doi: 10.1073/pnas.0702495104

Mumtaz, W., & Malik, A. S. (2018). A comparative study of different EEG reference choices for diagnosing unipolar depression. *Brain Topography*, *31*(5), 875-885

Muthukumaraswamy, S. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers In Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00138

Niedermeyer, E., & Lopes da Silva, F. (2010). *Electroencephalography* (6th ed., p. 1291). Philadelphia: Wolters Kluwer Health.

Noirhomme, Q., Brecheisen, R., Lesenfants, D., Antonopoulos, G., & Laureys, S. (2017). “Look at my classifier's result”: Disentangling unresponsive from (minimally) conscious patients. *Neuroimage*, *145*, 288-303. doi: 10.1016/j.neuroimage.2015.12.006

Ontario, H. (2004). Bispectral index monitor: an evidence-based analysis. *Ontario Health Technology Assessment Series*, *4*(9), 1-70. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387745/pdf/ohtas-04-70.pdf

Overgaard, M., & Overgaard, R. (2010). Neural Correlates of Contents and Levels of Consciousness. *Frontiers In Psychology*, *1*. doi: 10.3389/fpsyg.2010.00164

Overgaard, M., Rote, J., Mouridsen, K., & Ramsøy, T. (2006). Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Consciousness And Cognition*, *15*(4), 700-708. doi: 10.1016/j.concog.2006.04.002

Owen, A. (2008). Disorders of Consciousness. *Annals Of The New York Academy Of Sciences*, *1124*(1), 225-238. doi: 10.1196/annals.1440.013

Owen, A., Coleman, M., Boly, M., Davis, M., Laureys, S., & Pickard, J. (2006). Detecting Awareness in the Vegetative State. *Science*, *313*(5792), 1402-1402. doi: 10.1126/science.1130197

Pereira, A., & Ricke, H. (2009). What is Consciousness?: Towards a Preliminary Definition. *Journal Of Consciousness Studies*, *16*(5), 28-45.

Rautee, R., Sampson, T., Särkelä, M., Melto, S., Hovilehto, S., & Van Gils, M. (2004). Application of spectral entropy to EEG and facial EMG frequency bands for the assessment of level of sedation in ICU (pp. 3481-3484). San Francisco, CA, USA: The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Reuber, M., & Elger, C. (2003). Psychogenic nonepileptic seizures: review and update. *Epilepsy & Behavior*, *4*(3), 205-216. doi: 10.1016/s1525-5050(03)00104-5

Sanders, R., Tononi, G., Laureys, S., & Sleigh, J. (2012). Unresponsiveness ≠ Unconsciousness. *Anesthesiology*, *116*(4), 946-959. doi: 10.1097/aln.0b013e318249d0a7

Sarà, M., Pistoia, F., Pasqualetti, P., Sebastiano, F., Onorati, P., & Rossini, P. (2011). Functional Isolation Within the Cerebral Cortex in the Vegetative State. *Neurorehabilitation And Neural Repair*, *25*(1), 35-42. doi: 10.1177/1545968310378508

Sarasso, S., Boly, M., Napolitani, M., Gosseries, O., Charland-Verville, V., & Casarotto, S. et al. (2015). Consciousness and Complexity during Unresponsiveness Induced by Propofol, Xenon, and Ketamine. *Current Biology*, *25*(23), 3099-3105. doi: 10.1016/j.cub.2015.10.014

Sarasso, S., Rosanova, M., Casali, A., Casarotto, S., Fecchio, M., & Boly, M. et al. (2014). Quantifying Cortical EEG Responses to TMS in (Un)consciousness. *Clinical EEG And Neuroscience*, *45*(1), 40-49. doi: 10.1177/1550059413513723

Schartner, M., Carhart-Harris, R., Barrett, A., Seth, A., & Muthukumaraswamy, S. (2017). Increased spontaneous MEG signal diversity for psychoactive doses of ketamine, LSD and psilocybin. *Scientific Reports*, *7*(1). doi: 10.1038/srep46421

Schartner, M., Pigorini, A., Gibbs, S., Arnulfo, G., Sarasso, S., & Barnett, L. et al. (2017). Global and local complexity of intracranial EEG decreases during NREM sleep. *Neuroscience Of Consciousness*, *2017*(1), niw022. doi: 10.1093/nc/niw022

Schartner, M., Seth, A., Noirhomme, Q., Boly, M., Bruno, M., Laureys, S., & Barrett, A. (2015). Complexity of Multi-Dimensional Spontaneous EEG Decreases during Propofol Induced General Anaesthesia. *PLOS ONE*, *10*(8), e0133532. doi: 10.1371/journal.pone.0133532

Schiff, N., Nauvel, T., & Victor, J. (2014). Large-scale brain dynamics in disorders of consciousness. *Current Opinion In Neurobiology*, *25*, 7-14. doi: 10.1016/j.conb.2013.10.007

Schnakers, C., Vanhaudenhuyse, A., Giacino, J., Ventura, M., Boly, M., & Majerus, S. et al. (2009). Diagnostic accuracy of the vegetative and minimally conscious state: Clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology*, *9*(1), 35. doi: 10.1186/1471-2377-9-35

Scott, G., & Carhart-Harris, R. (2019). Psychedelics as a treatment for disorders of consciousness. *Neuroscience Of Consciousness*, *5*, 1-8. doi: 10.1093/nc/niz003

Seel, R., Sherer, M., Whyte, J., Katz, D., Giacino, J., & Rosenbaum, A. et al. (2010). Assessment Scales for Disorders of Consciousness: Evidence-Based Recommendations for Clinical Practice and Research. *Archives Of Physical Medicine And Rehabilitation*, *91*(12), 1795-1813. doi: 10.1016/j.apmr.2010.07.218

Seer Medical. n.d. Retrieved from https://www.seermedical.com/

Seth, A., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends In Cognitive Sciences*, *12*(8), 314-321. doi: 10.1016/j.tics.2008.04.008

Seth, A., Izhikevich, E., Reeke, G., & Edelman, G. (2006). Theories and measures of consciousness: An extended framework. *Proceedings Of The National Academy Of Sciences*, *103*(28), 10799-10804. doi: 10.1073/pnas.0604347103

Siclari, F., Baird, B., Perogamvros, L., Bernardi, G., LaRocque, J., & Riedner, B. et al. (2017). The neural correlates of dreaming. *Nature Neuroscience*, *20*(6), 872-878. doi: 10.1038/nn.4545

Silber, M., Ancoli-Israel, S., Bonnet, M., Chokroverty, S., Grigg-Damberger, M., & Hirshkowitz, M. et al. (2007). The visual scoring of sleep in adults. *Journal Of Clinical Sleep Medicine*, *3*(2), 121-131. Retrieved from http://www.aasmnet.org

Sitt, J., King, J., El Karoui, I., Rohaut, B., Faugeras, F., & Gramfort, A. et al. (2014). Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain*, *137*(8), 2258-2270. doi: 10.1093/brain/awu141

Siuly, S., Li, Y., & Zhang, Y. (2016). Electroencephalogram (EEG) and its background. *EEG signal analysis & Classification*. doi: 10.1007/978-3-319-47653-7-1

Sleigh, J., & Donovan, J. (1999). Comparison of bispectral index, 95% spectral edge frequency and approximate entropy of the EEG, with changes in heart rate variability during induction of general anaesthesia. *British Journal Of Anaesthesia*, *82*(5), 666-671. doi: 10.1093/bja/82.5.666

Tatum, W. (2014). *Handbook of EEG interpretation* (2nd ed.). Demos Medical Publishing.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*(1), 42. doi: 10.1186/1471-2202-5-42

Tononi, G. (2008). Consciousness as Integrated Information: a Provisional Manifesto. *The Biological Bulletin*, *215*(3), 216-242. doi: 10.2307/25470707

Tononi, G., Sporns, O., & Edelman, G. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proceedings Of The National Academy Of Sciences*, *91*(11), 5033-5037. doi: 10.1073/pnas.91.11.5033

van den Brink, R., Nieuwenhuis, S., van Boxtel, G., van Luijtelaar, G., Eilander, H., & Wijnen, V. (2018). Task-free spectral EEG dynamics track and predict patient recovery from severe acquired brain injury. *Neuroimage: Clinical*, *17*, 43-52. doi: 10.1016/j.nicl.2017.10.003

Vollenweider, F., & Kometer, M. (2010). The neurobiology of psychedelic drugs: implications for the treatment of mood disorders. *Nature Reviews Neuroscience*, *11*(9), 642-651. doi: 10.1038/nrn2884

Voss, L., & Sleigh, J. (2007). Monitoring consciousness: the current status of EEG-based depth of anaesthesia monitors. *Best Practice & Research Clinical Anaesthesiology*, *21*(3), 313-325. doi: 10.1016/j.bpa.2007.04.003

Wang, J., Noh, G., Choi, B., Ku, S., Joo, P., & Jung, W. et al. (2017). Suppressed neural complexity during ketamine- and propofol-induced unconsciousness. *Neuroscience Letters*, *653*, 320-325. doi: 10.1016/j.neulet.2017.05.045

Wassermann, E. (2000). Side effects of repetitive transcranial magnetic stimulation. *Depression And Anxiety*, *12*(3), 124-129. doi: 10.1002/1520-6394(2000)12:3<124::aid-da3>3.0.co;2-e

Zhang, X., Roy, R., & Jensen, E. (2001). EEG complexity as a measure of depth of anesthesia for patients. *IEEE Transactions On Biomedical Engineering*, *48*(12), 1424-1433. doi: 10.1109/10.966601

Appendix A

Coefficient of Variance (CoV) calculated for LZc during different Sleep/Wake categories (top), LZc scores obtained at different sample durations (centre), and LZc between different days of Wake (bottom). Participant 1 (left) and Participant 2 (right) shown separately.