A NEUROPHYSIOLOGICAL STUDY OF THE IMPACT OF MIND WANDERING DURING ONLINE LECTURES

by

Colin David Conrad

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
May 2019

Table of Contents

List of Tables

List of Figures

Abstract

Lecture-format Massive Open Online Courses (MOOCs) were once thought to have the potential to permanently change higher education, though they did not live up to their original hype. Among the many posited reasons for this failure is MOOCs' inability to effectively leverage multimedia to meet users' cognitive demands, such as the need to prevent mind wandering. One of the challenges with establishing the impact of mind wandering in the online classroom is the difficulty of measuring it, due to the complexity of the mind wandering phenomenon. In this dissertation, two research questions are explored. The first question concerns the neurophysiology of mind wandering and how it can be measured in the first place. Two studies are described which employ electroencephalography (EEG) to measure attention-related brain responses to auditory stimuli as participants sat through an e-learning video. The studies employed different measures of self-report, which yielded different neurophysiological responses. However, the responses were demonstrated to distinguish on-task and mind wandering states, suggesting an effective neurophysiological measure of mind wandering and other attention-related constructs. The second research question concerns the impact of mind wandering on the efficacy of online lectures. We employ multiple measures of mind wandering to investigate this impact and successfully identified the negative impact that mind wandering has on rote learning. Though we do not explore which teaching strategies are most effective, the results suggest that teaching strategies which limit mind wandering may be able to improve rote learning outcomes. This leaves open possibilities for future research which explore such strategies for improving MOOC technology, but also for measuring attention-related information systems constructs.

Glossary

active learning Teaching techniques that find ways of engaging students' interest or attention by making them active in the learning practice.

alerting The brain network originally identified by observing sustained vigilance and sensitivity to attention stimuli.

attention The phenomenon of focusing on a single stimulus out of several simultaneously possible objects or trains of thought. The phenomenon is the result of at least three neuralogical processes: alerting, orienting and executive control.

Attention Network Test (ANT) A test battery developed by Fan et al. (2002) that uses a combination of flanker tests, which are designed to measure response inhibition, as well as cuing tasks demonstrated to measure attentional orienting.

brain-computer intferface (BCI) A computer program that takes direction from a direct communication pathway to the brain. In this dissertation, we describe BCIs which are powered by EEG signals.

cognitive absorption An influential construct proposed by Agarwal and Karahana (2000) which measures a state of deep engagement with an information technology.

cognitive load A construct used to describe effort being used in working memory.

Cognitive Theory of Multimedia Learning (CTML) Cognitive Theory of Multimedia Learning. Advanced by Mayer over the course of a long career, this theory states that "people learn more deeply from words and pictures than from words alone".

covert orienting General attentional orienting unrelated to a particular space.

decoding The process of identifying a relationship between a particular neural signal and a mental state.

eeg waves Patterns of central nervous oscillatory activity at particular band frequencies.

e-learning Learning technology that incorporates multimedia and information technology to deliver a learning experience.

electrodermal acivity (EDA) Electrophysiological measurement of the skin to measure activity of the central nervous system.

electroencephalography (EEG) Electrophysiological mesurement of the scalp to measure activity of the brain.

event-related potential (ERP) An electrical pattern evoked by a stimulus resulting from neural activity.

executive control The attention network originally conceptualized by Posner and Petersen (1990) to describe target detection and the limited capacity of attention.

eye-fixation related potential (EFRP) A pattern of eye movement fixations that are correlated with the P3 event-related potential.

flow A state under which individuals are intrinsically motivated to engage in an activity and in a state of deep attention to it.

functional magnetic resonance imaging (fMRI) A machine that measures brain activity by detecting changes in neuron blood flow. Usually leveraging the blood-oxygen-level dependent (BOLD), fMRI can be considered a secondary physiological measure of brain activity.

independent component analysis (ICA) An unsupervised machine learning method for separating a multibariate signal into statistically distinct components. In this dissertation, the technique is used to separate electrical noise (eg. blinks) from electrical activity generated by the brain.

Information Systems (IS) Refers to "socio-technical systems comprising of individuals or organizations and their interaction with business information technology." (Recker, 2012) We usually refer to Information Systems as it is practiced in business schools.

linear discriminant analysis (LDA) A method used in statistics and machine learning to find a linear combination of features that separates two or more objects. LDA is commonly used to reduce the dimensionality of data.

machine learning Statistical and computational techniques that give computers the ability to progressively improve performance on tasks without being explicitly programmed.

magnetoencephalography (MEG) A neuroimaging technique that maps neural activity by observing magnetic fields produced in the brain.

mind wandering Self-directed thoughts about a subject other than the primary task the participant is supposed to be engaged in.

MOOC Massive Open Online Course. Usually, though not always, delivered in a lecture format.

NASA Task Load Index (TLX) A widely-used questionnaire scale created by Hart et al. (1988) to measure perceived workload in a given task.

NeuroIS Neuro-Information Systems. The sub-discipline of Information Systems that use methods from cognitive and affective neuroscience to answer questions related to information technology and its use.

oddball paradigm An experimental design involving the presentation of repeated stimuli where an infrequent number of stimuli are different from the majority.

orienting The brain network that governs the mechanism of feature selection.

oscillatory activity Repetitive patterns of neural oscillations that can be observed using EEG. Neural oscillations are often measured in frequency bands (eg. alpha (8-13 Hz), beta (13-30 Hz), etc.).

overt orienting Attentional orienting toward a fixed point in space.

P1-N1-P2 complex A series of even-related potentials triggered by early attentional control mechanisms. This complex is detected using EEG or MEG and is sensitive to both visual and auditory stimuli.

P3 component An attention-related positive event-related potential observed at the 300 ms mark.

positron emission tomography (PET) A machine that observes metabolic processes in the body, in our case the brain, by tracing a radioactive tracer in the blood stream.

rote learning The ability to memorize and recall something taught.

STEM Science, Technology, Engineering and Mathematics university programs.

support vector machine (SVM) A supervised machine learning algorithm that classifies data according to a non-probabilistic kernal classifier that best separates points on a hyperplane. This technique is primarily used in highly dimensional datasets.

Acknowledgements

It is sometimes said that it takes a village to raise a child; it also also seems to take a village to raise a scholar. This work is the culmination not just of my PhD, but also the guidance and efforts of four academic communities at Dalhousie University. The Dalhousie Philosophy department introduced me to scholarship. Dalhousie's Faculty of Computer Science opened a world of new capabilities and possibilities. The Faculty of Management taught me which questions to ask. The Department of Psychology and Neuroscience taught me how to answer them. To everyone who are members to these communities or of the Dalhousie interdisciplinary community broadly, thank you for your support over the years. A special thank you to Lynne Robinson, the Interdisciplinary PhD Program Director, for being a champion of the big picture!

There are three people I would like to especially acknowledge who have had an exceptional positive influence over this work. Though Vlado Keselj is not formally recognized as a supervisor of this dissertation, he functioned as the supervisor of the computer science portion of my PhD and has been a longtime supporter of this work. Michael Bliemel introduced me to the Information Systems community his wisdom, guidance and passion for education has been instrumental to delivering this dissertation. I would like to acknowledge Aaron Newman whose patience, enthusiasm for innovation, and strong scholarly values have guided me throughout this journey. Without this supportive supervisory team I would not have been able to navigate the sometimes murky waters between the disciplines. Also, thanks go to to Rita Orji for opting to be part of this committee at the last minute!

I acknowledge the financial support of the Nova Scotia Graduate Scholarship, the Faculty of Computer Science, the Faculty of Management, the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Killam Trusts for supporting me financially.

I would like to acknowledge the role of family and friends throughout my PhD. My partner Jenny provided me ongoing love, support and enthusiasm for which I am very grateful. I acknowledge my close friends John Doucette and Jake Summers for their consistent willingness to hash through research ideas; realistically this has been going on our entire adult lives. I would like to thank my parents Janet, David and my brother Drew; I lack the space to properly thank you for the many years of love and support. Finally, though I cannot hope to comprehensively thank acknowledge all colleagues and groups who have supported me through the PhD, I would like to acknowledge some by name: Hossam Ali-Hassan, Benoit Aubert, Lyam Bailey, Shea Balish, Louis Beaubien, Katie Douglas, Cindy Hamon-Hill, Magdalena Jankowska, Mary Kilfoil, Jakie Kinley, Dijana Kosmajac, Ed Leach, Farrukh Momin, Stacey Taylor, Francesco Usai, Aarnoud Van Der Spoel, All Nations Church, the Global Shapers Community, the IDPhD Public Scholars, the NeuroIS Retreat, the Norman Newman Centre for Entrepreneurship, Shiftkey Labs, the Surge Science Innovation Sandbox and many others. Thank you all for your support through this life journey.

Chapter 1

Introduction

In 2011, then Stanford professors Sebastian Thurn and David Evans launched a free online course called *Introduction to Artificial Intelligence*. The course consisted of recordings of their usual online robotics lectures, but was given away publicly, for free. The initiative was wildly successful, attracting over 160 000 students from across the globe (Hsu, 2012). The apparent success inspired Thurn to found Udacity, one of the market leaders in the Massive Open Online Course (MOOC) movement. In their early days, MOOCs were thought to have the potential to permanently transform the institution of higher education by making top education content available to everyone in the world for free.

As time passed however, it quickly became clear that MOOCs would not succeed at transforming higher education as Thurn had originally envisioned. The wild success turned to cautious skepticism as it became clear to educators that MOOCs were lacking something. A 2013 study of a million users of University of Pennsylvania courses offered online through Coursera, one of Udacity's nonprofit competitors, found that only five percent of participants actually finished their classes (Perna et al., 2013). Other studies soon found similar results (Konnikova, 2014). More worringly, studies found that the students who were most likely to complete MOOCs were students who already had advanced degrees (Kolowich, 2013). Today, it is largely agreed that MOOCs have not lived up to the original expectations of Thurn and the hype of 2012. Writing in 2019, it is easy to form an opinion that MOOCs succeeded being an excellent supplementary learning tool but largely failed to instill the ability to learn characteristic of a university culture.

Though there are probably many reasons, one possible reason why MOOCs did not succeed is that they did not incorporate the right techniques in their design. Robert Ubell, for instance, argued in IEEE Spectrum that the downfall of MOOCs was driven by the fact that they did not incorporate active learning techniques (Ubell, 2017).

Ubell cites a meta analysis of 225 studies which found undergraduate Science Technology Engineering and Mathematics (STEM) classrooms employing active learning saw about 6% higher test scores and were considerably less likely to have students fail the course (Freeman et al., 2014). Active learning is often identified as the theory behind pedagogical techniques that engage students' interest or attention by making them active in the learning practice (Young, Robinson, & Alberts, 2009). As Ubell's reasoning goes, the ability to use active learning is one of the underlying reasons why high quality in-person lectures at universities work better than pre-recorded MOOCs. The result is an in-person experience that can adapt to students' changing attention states; something that MOOCs cannot easily replicate.

Similarly, other scholars have argued that the inability for online lectures to capture and maintain attention is also to blame, particularly for their inability to prevent mind wandering (Szpunar, Moulton, & Schacter, 2013). Mind wandering describes a shift in attention away from the primary task and towards unrelated self-generated thoughts (Smallwood & Schooler, 2006). Mind wandering has been demonstrated to impact performance during monotonous tasks such as driving long distances (Y. Zhang & Kumada, 2017). It follows that students who experience mind wandering during long lectures would similarly learn less, as they are engaged with their own self-directed thoughts rather than the lecture content. This led some education scholars to conclude that techniques designed to minimize mind wandering during lectures would have a positive impact on student retention (Smallwood & Schooler, 2015). Their reasoning likewise holds that one of the failures of MOOCs is their inability to incorporate such techniques; many MOOC users seem to simply lose attention.

Can we really be sure that MOOCs' inability to actively involve students or capture their attention is to blame? It seems intuitive to professors that this is the case because we had first-hand experience of mind wandering during lectures, and likely had personal experience of missing out on relevant lesson content at some point. Yet, many of us can also recall having positive learning experiences from mind wandering. For many, including the author of this dissertation, the mind wandering experience has yielded moments of creativity or deep analytical insight into the lecture content being delivered. There is an argument to be made that mind wandering or attention diversion may not be detrimental to education at all, and that mind wandering may indeed play a crucial role in the attainment of a deep, critical education. It would follow, then, that the inability to capture attention is not one of the causes of MOOCs' failure, but something else entirely.

This dissertation takes steps towards answering the question of whether user attention and mind wandering has an impact on MOOC learning outcomes. The argument being advanced is that the loss of student attention has a largely negative relationship with rote learning and also that a person's ability to pay attention plays a crucial role in the efficacy of e-learning technology. The relationship between attention and e-learning efficacy has only recently come to focus, possibly because we did not have a good way to measure attention in the past. Much of the work on attention and e-learning research (or in Information Systems broadly) has focused on behaviour or self report, which is intrinsically problematic for this research problem. We instead take a neurophysiological approach to explain the relationship.

1.1 Why Observe Neurophysiology in Information Systems Research?

The subject of Information Systems (IS) refers to the "socio-technical systems comprising of individuals or organizations and their interaction with business information technology," of which online learning technology clearly plays a part (Recker, 2012). Though this dissertation explores specific observations about humans' interaction with MOOCs, a specific type of learning information system, it is also part of a broader movement in the Information Systems field to incorporate the findings and techniques of Neuroscience. This community of Neuro-Information-Systems (often stylized "NeuroIS") researchers seek to extend the enterprise of Information Systems by incorporating Neuroscience as a reference discipline. The research in this dissertation shares a number of motivations with the wider NeuroIS endeavour which should be explained.

Many of the techniques employed in the discipline of Information Systems has its roots in Psychology. Psychology research has been reliably conducted for nearly a century, and has created theories about the relationships between mental states and behaviours. Many of the greatest contributions in the Information Systems literature has likewise utilized these techniques to create theories that have explanatory power. However, though they are reliable, behavioural theories tell us little about the antecedents of behaviours, particularly the antecedents that participants are not consciously aware of. Riedl and Léger make this argument well, arguing that Neuroscience tools can help overcome these limitations by accurately measuring the factors that are the antecedents to behaviours, often with higher rates of accuracy than psychometric or behavioural methods (Riedl & Léger, 2016). As we will see later in the dissertation, the brain correlates are particularly relevant to attention-related constructs. Findings in neuroscience have discovered that attention consists of at least three distinct processes, rather than one single intuitive process. This can offer great insight into theories about the role of attention in information technology use.

A second motivation for our use of neurophysiological measures is practical. The indicators that we employ give us a series of objective observations in real time. Changes in attention processes are often gradual, and not easily consciously understood by participants. The constructs that draw from attention processes (such as mind wandering) are likewise best observed in real time. The ability to measure in real time also has implications for the design of better information systems (Brocke, Riedl, & Léger, 2013; Dimoka et al., 2012). Real time measures further open the possibility of a passive brain-computer intferface (BCI), where user experiences change depending on the users' brain signals. By employing machine learning with the neurophysiological data, we can create reliable adaptive systems that change user experience without disrupting the user (Randolph et al., 2015). This factor is critical for certain applications where distraction from the task is likely to disrupt the phenomenon that we are measuring.

There is also a third motivation which has so far been largely ignored by the academic information systems community, but is commonly discussed by a number of prominent business leaders. There is a growing awareness about the transformative impact of the recent generation of technologies and their impact on human society. Many of these technologies promise to result in entirely new information systems. Neurotechnologies have been specifically identified as one of the technologies with the potential to drive fundamental social change in the near future. For instance, Klaus Schwab, the founder and Executive Chairman of the World Economic Forum wrote (Schwab, 2017, p. 98):

> The mind-boggling innovations triggered by the fourth industrial revolution, from biotechnology to AI, are redefining what it means to be human. They are pushing the current thresholds of life span, health, cognition and capabilities in ways that were previously the preserve of science function. As knowledge and discoveries in these fields progress, our focus and commitment to having ongoing moral and ethical decisions is critical.

Schwab calls this nexus of advances in biotechnology and AI the foundation of a "Foruth Industrial Revolution" which is currently transforming how businesses and institutions operate. Concerning neurotechologies specifically, he cites the potential for neurotechnologies to extend or improve human cognitive abilities as one of the potential transformative impacts (Schwab, 2017, p. 171). Clearly, the potential emergence of new information systems based on neurotechnologies is a motivation for studying the potential ways such systems may emerge.

In short, the primary motivations for embarking on a neurophysiological approach to this research subject of MOOC efficacy is that it offers new research methods to tackle difficult questions, and may form the technical foundation new information systems in the future. To the author of this dissertation, neuroimaging is fundamentally a tool which is instrumental to pursuing information systems research. Though much of this dissertation will focus on constructing new measures that draw heavily on material from cognitive neuroscience and computer science, it is important to keep in mind that the construction of these measures is not the ultimate goal of the dissertation. By leveraging insights from cognitive neuroscience, as well as from machine learning, we can expand the horizons of the Information Systems discipline.

1.2 Thesis Statement and Research Contributions

This all said, the task before us is to conduct interdisciplinary research that draws from the fields of Education, Neuroscience, Information Systems, Machine Learning and to a lesser extent, Philosophy. Rather than solving the broad question of how MOOCs should be optimized to meet users' attentional needs, we focus on one particular attention-related construct, mind wandering, and whether neurophysiological measures contribute to its measurement. The first fundamental question that this thesis addresses can be summarized as:

RQ1 Can neurophysiological measures be used to detect differences in mind wandering during online lectures?

Fortunately, we are not the first to explore this question, as Braboszcz and Delorme discovered such correlates in a meditation study (Braboszcz & Delorme, 2011) and Smallwood et al. observed correlates during a sustained attention response task (Smallwood, Beach, Schooler, & Handy, 2008). However, it is unclear whether these measures can translate into the MOOC context, as users are presented with audio and visual stimuli which could confound the results. In Chapter 4, we describe an experiment that validates these electroencephalography measures in a MOOC setting. In the experiment, participants were asked to participate in a MOOC lesson as auditory signals were administered by the computer. Participants were also asked to report when they experienced mind wandering. Differences participant brain responses to auditory signals before and after the mind wandering response resulted in a measure that can reliably distinguish between the two reported states. The experiment then explores how machine learning can be used to create a real-time neurophysiological measure that can classify these brain responses in real-time. Recognizing that measurement is only the first step in our motivation, we can articulate the second questions as follows:

RQ2 Does mind wandering affect how well we learn from online lectures?

To the best of our knowledge, we are the first to explore this in the context of learning with a focus on mind wandering and the neurophysiology of attention using ERP. In order to answer this question, we must also dive into the relationship between the neurophysiological measures and those from more conventional Information Systems research. Chapter 5 describes an experiment where participants again participated in a MOOC lesson as auditory signals were delivered, but were instead probed for experienced mind wandering at various intervals throughout the lesson. Participants were asked to complete tests on the MOOC topic before the lesson began, and were asked to complete the same quiz afterwards. They were also asked to complete questionnaires on their perceived mind wandering and the related task load. In this study, we use psychological questionnaires that have been used to measure mind wandering and task load in the information systems context (Sullivan, Davis, & Koh, 2015; Hart &

Staveland, 1988). We cross validate these with both learning outcome acquisition and neurophysiological measures in order to explore both the research value of attention neurophysiology in IS and the role of mind wandering on MOOC efficacy.

This rest of this dissertation is structured as follows. In Chapter 2 we discuss background literature from the four subject domains in the context of online lectures. We begin that chapter by exploring dominant theories of learning, before exploring the neurophysiology of attention, its relevance to information systems, and the technical methods employed in this work. In Chapter 3, the detailed hypotheses and research methodology used in this dissertation are discussed. Chapter 4 discusses the first experiment, which establishes a real-time attention measure for MOOCs and other e-learning applications. Chapter 5 describes the second experiment, which explores the relationship between the neurophysiological measures, different psychometric measures and the attainment of learning outcomes. We then discuss the implications of this research and conclude in Chapter 6, ultimately ending with a focus on future research questions and the potential applications of this work to other IS domain problems.

Chapter 2

Background and Related Work

Given that the task of this dissertation is to discover how neurophysiological indicators of attention can lend new insight into e-learning efficacy, the research described in this dissertation is fundamentally interdisciplinary. It draws on material from four disciplines: Education, Neuroscience, Information Systems and Machine Learning. Before proceeding, it will be helpful for readers to review the relevant material from the various disciplines which this work draws from. Rather than reviewing the disciplines at length, this chapter discusses the relevant material with a particular focus on how they inform the research in question. We will subsequently describe many of the subjects mentioned in more detail, but we provide a broad discussion of the subjects here. We take a narrative approach to this review, which not only makes the material more readable, but helps drive the research motivation and methods described in Chapter 3.

2.1 What Does Learning Theory Say About Online Lectures?

There has been considerable recent academic interest focused on the factors that affect MOOC completion. In fact, in the journal Computers & Education alone, 36 papers were published with the keyword "MOOC" or "MOOCs" in the first half of 2018. When contrasted with 17 papers in in the entire year of 2017, we can confidently say that there is a growing interest in the issue. Much of this research has revealed a consistent story. The greatest factors behind student completion of MOOCs appears to be the degree of interaction with instructors and other learners (Hone & El Said, 2016; Gregori, Zhang, Galván-Fernández, & de Asís Fernández-Navarro, 2018; Jung & Lee, 2018), as well as the students' motivation for learning from the MOOC in the first place (Zhou, 2016; J. Zhang, 2016; Phan, McNeil, & Robin, 2016; Haslam et al., 2019). The findings have overwhelmingly demonstrated that MOOCs with high completion and success rates 1) have a higher degree of instructor to student interaction and 2) are taken by students with the expectation of some tangible credit outcome. These have significant implications for the design of effective MOOCs and their most effective uses.

These studies have focused largely on the nature of the study participants rather than the design of the MOOC content. We will instead focus on the interaction between individual participants and the content of MOOCs. Much of the work on this subject predates the MOOC boom. As mentioned in the introduction, e-learning is by no means a new subject, and has been a topic of interest to researchers since at least the 1980s. In fact, famous MOOCs such as Udacity incorporate many of the best practices developed during the early days of e-learning, in addition to findings from recent research. In this section, we will explore four theories that have been particularly influential on the advancement of e-learning and content design. We shall see that the subject of attention and its measure is an important factor in all of them.

2.1.1 The Cognitive Theory of Multimedia Learning

Perhaps the most cited theory of successful e-learning is the Cognitive Theory of Multimedia Learning (CTML), which lays the foundation for many of the widely cited best practices in online learning (Clark, Mayer, & Thalheimer, 2003). Advanced by Mayer over the course of a long career, this theory broadly states that "people learn more deeply from words and pictures than from words alone" (Mayer, 2009). The reasoning makes intuitive sense and may seem unremarkable at first glance to an uncritical reader. However, this theory goes a long way to explaining the potential of rich multimedia for improving the learning experience, and for how a good MOOC could be designed. The CTML is built on three assumptions which will be explained in turn.

The first assumption is what Mayer calls the *dual channel* assumption. This assumption is that "humans possess separate information-processing channels for visually represented material and auditorially represented material" (Mayer, 2009, p. 64). This assumption explains how multiple media (i.e. words plus pictures) can affect people differently than a single medium (i.e. just words). If we have different mental or physical processes for different types of media, we can see how the interplay between the processes matter. Though, as Mayer rightly points out, the channels can interact in the processing of abstract concepts, the channels investigate stimuli different types of experiences separately.

The Dual Channel assumption is rooted in the work of Paivio, and his dual-coding approach (Paivio, 2014). The dual-coding approach is a theory about how our brains handle mental representations. It finds a distinction between how representations are handled by verbal and non-verbal systems and argues that they are ultimately coordinated by different brain networks. When related to Mayer's thesis, this helps explain not just the difference between auditory and visual stimuli, but also between written words and pictures. According to Mayer, the result is that words and pictures are processed by distinct systems, regardless of whether the words are written or said. It likewise follows that there could be other distinct systems that are relevant the dual-channel assumption, but they will not be discussed in this scope of multimedia.

The second assumption made by Mayer is the *limited capacity* assumption. This assumption stats that "humans are limited in the amount of information that can be processed in each channel at one time" (Mayer, 2009, p. 66). It explains how humans can't simply absorb information at an infinite capacity. Together with the dual-channel assumption, it follows that it would be productive to break down learning content into material for different channels, so that any one channel does not become overloaded.

This assumption is rooted in Baddley's conception of working memory and Sweller's theory of cognitive load (A. D. Baddeley & Hitch, 1974; A. Baddeley, 1992; Sweller, 1988, 1994). Working memory is an intuitive concept, and was perhaps most best demonstrated though a digit span test. Miller famously demonstrated that healthy people cannot retain more than five to seven digits at a time (Miller, 1956). Though you can train yourself to develop some sort of strategy to remember chunks of information (i.e. remember "56" instead of "5" and "6"), Miller showed that there are limits to working memory. Baddeley's concept of working memory roots it in the brains' central executive system, which controls the allocation of cognitive resources.

Sweller's theory of cognitive load takes this further. Where working memory is responsible for processing and retrieving task-relevant information, Sweller's model asserts that the use of working memory also plays a role in the acquisition of new knowledge (Sweller, 1994). According to Sweller, when working memory is engaged in effort to understand a concept, the concept is eventually abstracted into a *schema*. More complex ideas can in turn be assessed using memory and further abstracted, which explains how education works. Too much strain on working memory will likewise prevent abstraction from taking place.

Sweller later distinguishes *intrinsic*, *extraneous* and *germane* cognitive load; three types of cognitive load that each play different roles in learning (Sweller, Van Merrienboer, & Paas, 1998). Intrinsic cognitive load describes the absolute difficulty of an activity, while extraneous cognitive load describes the mental effort dedicated to the instruction presentation, and germane load describes the effort required to abstract a concept into a schema. This later distinction helps further inform the Cognitive Theory of Multimedia Learning (CTML) by lending support for the careful design of the multimedia. By keeping the stimuli simple and not unnecessarily rich, it follows that extraneous load is limited, devoting more resources for the process of schema formation.

The third assumption described by Mayer is the *active processing* assumption. This assumption states that "humans actively engage in cognitive processing in order to construct a coherent mental representation of their experiences" (Mayer, 2009, p. 67). Like Sweller's concept of a *schema*, the active processing assumption holds that humans are naturally in the business of creating abstractions. Unlike Sweller's theory, the Active Processing assumption holds that abstraction is a series of processes rather than one process limited by the central executive system.

Mayer also holds that there are three processes essential for active learning: selecting relevant material, organizing selected material, and integrating selected material with existing knowledge. The first of these is fundamentally an attention process. Selecting relevant material "occurs when a learner pays attention to appropriate words and images" in the material currently being digested into the central executive system (Mayer, 2009, p. 70). When the learner switches to organizing the material, she or he then arranges the presented material conceptually before making connections and integrating it with past experience. The result is new knowledge.

Though all three of the CTML's assumptions are rooted in well studied phenomena in the learning literature, the final active processing assumption is particularly interesting because the relationship between active learning assumption and learner attention is not explicitly stated by Mayer (2009). Though it makes sense that learners *should* be actively engaged in order to learn, it is clear that not all learners are *in fact* attending to a given lesson. The CTML merely assumes that they are. If the CTML alone was sufficient to predict learning success, MOOCs would not suffer from the problems of low retention and completion.

In later work, Mayer raises limitations to the CTML and builds on this critique (Mayer, 2014). Though there is evidence that effective learning requires the limitation of extraneous cognitive load, he mentions that the CTML can be extended by investigating principles that foster motivation and generative processing. One such way the CTML could be complemented is with teaching techniques designed to engage, capture or sustain attention, or attention-related phenomena. As pedagogues, it is an age old problem. Engaging all students in a given lesson is very difficult and perhaps is a problem that is increasingly prevalent in a world where multimedia is ubiquitous. Attention therefore offers a potential area of enquiry for extending e-learning best practices established by the CTML.

2.1.2 Active Learning and Student Engagement

Another prominent theory that can be extended by a robust study of attention is active learning. A great deal of proverbial ink has been split on developing teaching techniques that involve students in novel ways, and not all of it has been written in the shadow of Mayer's seminal work. The concept of active learning, has been particularly influential in the movement to innovate in the classroom. Active learning theories build on similar reasoning as Mayer's active processing assumption, and describe techniques to find ways of engaging students' interest or attention by making them active in the learning process (Young et al., 2009). For instance, Young et al. describe *vigilance decrement*, a gradual decrease in student vigilance during lectures that exceed 10 to 30 minutes. They find that the loss of vigilance negatively impacts learning outcomes. They also make recommendations about how to combat the vigilance decrement using multimedia, as well as interactive feedback or in-class group work.

This is what motivates criticism about the design of MOOCs from the perspective of active learning. As discussed in the introduction, Ubell's criticism of lecture-style MOOCs is that they not only failed to incorporate best practices in their design, but that they failed to use the full potential of multimedia to engage students (Ubell, 2017). Freeman et al. conducted a meta analysis of the active learning literature and found that classrooms which employed active learning saw on average 6% higher test scores and had 50% higher completion rates (Freeman et al., 2014). In that study, they defined active learning activities to include cooperative group activities, clickers, problem-based learning, in-class worksheets and studio classrooms. For the purposes of their study, papers which included at least one of these techniques for during 10% and 100% of the class time were considered to be using active learning techniques. They were contrasted with traditional lectures and participants in courses which incorporated active learning lectures were found to perform better on examinations. These findings will not be surprising to experienced university teachers, who have likely experienced student disengagement during lectures firsthand.

In business and technology classrooms, active learning is particularly important and is employed to teach concepts that are either intrinsically uninteresting to their students, or technically complex. For instance, big data infrastructure has been recognized to be crucially important for Management Information Systems students, but is quite detached from the regular business school curriculum (Phillips-Wren, Iyer, Kulkarni, & Ariyachandra, 2015). An alternative approach to traditional lecture could be to use hands-on analogy, which employs a number of the aforementioned techniques to overcome technical barriers while helping maintain student attention. Analogy has been implemented in the teaching of science to facilitate inferential learning of complex concepts (Niebert, Marsch, & Treagust, 2012; Treagust & Duit, 2015). For instance, when teaching the concept of cell division to high school students, a teacher can use the analogy of breaking a piece of chocolate to represent how cells can be divided. The analogy of chocolate is an interactive, if imperfect analogy. Like cells, the chocolate bar can be divided multiple times into many pieces. Unlike cells, the pieces of chocolate become smaller. Such an activity interactively illustrates a concept similar to the one being discussed, but the nominal flaws with the analogy foster meaningful discussion on the concept.

In a forthcoming paper, we explore the application of hands-on analogy to the

Management Information Systems classroom (Conrad, Bliemel, & Ali-Hassan, 2019). The paper describes *Hadoop Hands On*, a technique for teaching students about the MapReduce algorithm using playing cards. MapReduce is an algorithm for performing distributed computer processing tasks through a computer cluster, and has played a key role in the expansion of big data. Understanding the technical details of Hadoop requires knowledge of computer science, which most management students lack. By using analogy, students were able to form abstractions about the key concepts without having technical knowledge. By having an interactive component, the classroom was able to facilitate learning by drawing students into the exercise. Drawing students in this way theoretically better captures their attention and fosters reflection on the teaching material. The study also revealed that a key contribution to students' learning success was the study's ability to trigger *flow*, another attention-related construct. We shall explore flow in detail, as it has played a role in a number of innovations in interactive multimedia learning and is related to the central argument of this dissertation.

2.1.3 Flow and the Facilitation of Active Processing

The concept of flow was originally proposed by Csikszentmihalyi in the early 1990's, and has since been influential in various disciplines ranging from online learning to information systems to youth sport psychology. Flow, in its original conception, described a state under which individuals are intrinsically motivated to engage in an activity and in a state of deep attention to it (Csikszentmihalyi, 2014). Csikszentmihalyi suggests that the flow state is fundamentally *enjoyable*, but also constitutes a deep absorption in an activity. Flow can be triggered by activities such as "play, art, pageantry, ritual and sports" but also by the play of ideas, philosophy and the acquisition of new knowledge. In this final respect, flow could be experienced in the conditions that make learning possible. For those of us who have taken up higher learning as a vocation, we may deeply relate to this notion, as many of us find the act of learning to be intrinsically motivating. In this way, it is easy to see the potential interplay between flow and the active processing assumption; students are more likely to actively process information if they are intrinsically motivated to stay on task.

Past work on measuring flow in learning situations is varied. Conventional work on the subject involves creating psychometric scales to measure participant perceptions of flow following an experience, and studying its relationship to the attainment of learning outcomes (Jackson & Eklund, 2004; Ullén et al., 2012). Perhaps most notable, Jackson et al. (2004) published the *The Flow Scales Manual* which is a collection of psychometric scales used to capture flow. In the Information Systems literature, the cognitive absorption scale developed by Agarwal and Karahana has been extensively utilized to observe the role of the cognitive absorption construct and technology use (Agarwal & Karahanna, 2000). The Cognitive Absorption construct is closely related to flow, and is rooted in Csikszentmihalyi's theory.

Flow and cognitive absorption have been used to study the effectiveness of e-learning technologies, and especially the category of technologies referred to as "serious games". Serious games are games used to teach a concept, and usually take the form of interactive video games that leverage simulation. These technologies have been observed to have a positive impact on the classroom learning environment (Lu, Hallinger, & Showanasai, 2014), and their ability to trigger flow or cognitive absorption is often claimed to be one of the reasons for this (Léger, Davis, Perret, & Dunaway, 2010). The reasoning behind this is that they draw users into an engaging experience that triggers flow, and in doing so reflect a higher degree of engagement and active processing.

However, there is a major limitation with measuring flow using conventional scales administered following an extended experience. Flow is an state that happens throughout a given period of time, and a person can expect to enter and leave the flow state multiple times throughout that observed period. Hoffman and Novak attempted to measure flow for the purposes of e-commerce and online marketing, but it proved to be too elusive to be useful (Hoffman & Novak, 2009). Research which utilizes the post-hoc flow scales are severely limited in what they measure, as they measure people's post-hoc perceptions of experienced flow rather than the actual flow experience.

A potential solution to this is to employ a real-time questionnaire measure. Pearce has suggested one such measure, specifically in the context of online learning (Pearce, Ainley, & Howard, 2005). The Pearce measure consists of a simple ratio of perceived challenge and skill, inspired by Csikszentmihalyi's original conception of flow.

However, this measure remains subjective, relying on participant's perceived balance between challenge and skill. Perhaps more worrying is that the measure is disruptive and requires researchers to remove people from the flow state in order to measure it.

A second category of research on this subject involves identifying psychophysiological correlates of the flow state and measuring these when the flow state is expected to be triggered. Measures that leverage tools such as electroencephalography (EEG) and electrodermal acivity (EDA) have been found to be correlated in both the classroom learning and e-learning learning contexts (De Manzano, Theorell, Harmat, & Ullén, 2010; Léger, Davis, Cronan, & Perret, 2014). Léger et al. (2014), who explored ways to measure the flow state using electroencephalography, is of particular interest because they also measure the correlates of cognitive absorption in the context of a serious game. By observing patterns of neural oscillations and electrodermal activity, they establish correlates of cognitive absorption which can overcome the limitations described by Hoffman and Novak (2009). Using psychophysiological measures they were able to measure correlates objectively and better understand the constituent processes that comprise the flow construct. Léger et al. (2014) ultimately found correlations between alpha band activity and flow, which we shall see is also correlated with attentional processes. A robust study of attention therefore has the potential of giving a deeper and more specific understanding of how constructs such as flow facilitate successful learning.

2.1.4 The Role of Mind Wandering in the Classroom

Another parallel line of inquiry has been the work on the relationship between mind wandering, on-task thought and successful learning. As we saw in the introduction, this has been explored in the context of learning before. Mind wandering refers to a series of processes commonly referred to as "daydreaming," or "spontaneous thoughts", and has been proposed to play a role in human cognition, particularly in the routine consolidation of past experiences (Christoff, Gordon, Smith, & Vancouver, 2011). Mind wandering has been of recent interest to the academic psychology community (Drescher, Van den Bussche, & Desender, 2018; Gonçalves, Carvalho, Mendes, Leite, & Boggio, 2018) and has also been observed in the applied context of driving, where it is demonstrated to have a negative impact on performance (Baldwin et al., 2017; Y. Zhang & Kumada, 2017). At first glance, it would seem obvious that mind wandering contrasts with active processing related to learning content. After all, attending a lecture seems to be a sustained attention task. When students' minds are wandering, attention away from a lesson or task and toward something else.

There is considerable research which suggests that mind wandering does indeed occur in the higher education classroom environment, and that it indeed has an overall negative impact on student performance. Lindquist and McLean examined the impact of daydreaming and task-unrelated thoughts on the success of learning outcomes (Lindquist & McLean, 2011). In that study the authors used a combination of self report and experience samples during a lecture, and compared the measures to academic performance at the end of term. They found significant negative correlations between task-unrelated thoughts and academic performance. It follows to reason that mind wandering can be disruptive to education and should be limited, which is a conclusion similarly shared by experts on mind wandering, such as Smallwood and Schooler (2015). However, though recent research has reaffirmed that mind wandering has a negative impact on student performance, it has also established that our intuitions about the ubiquity of mind wandering may be incorrect (Wammes, Boucher, Seli, Cheyne, & Smilek, 2016; Wammes, Seli, Cheyne, Boucher, & Smilek, 2016; Wammes & Smilek, 2017). Wammes et al. (2016a) found that rates of unintentional mind wandering during live lectures were low at 14% of probes and that mind wandering did not increase as the lecture went on.

There is also evidence that performance during online lectures is worse than live lectures and that this may be due to mind wandering. In 2012, San Jose State University launched a partnership program with Udacity where students were offered the opportunity to either take introductory mathematics or statistics courses live or using Udacityś platform. Students who took the course for credit and used Udacity's platform were found to be significantly less likely to complete or pass the course (Firmin et al., 2014). Wammes and Smilek (2017) later investigated the differences between in-person lectures and their online counterparts and found that online lectures follow a different trend than live lectures, with participants reporting higher degrees of mind wandering as the lecture went on. Using a five-point Likert to measure degrees of mind wandering, they compared groups enrolled in a live version of a lecture with those enrolled in an online video. They found that participants enrolled in the video were significantly more likely to report heightened degrees of mind wandering later in the video (after 51 minutes) but those enrolled in the live lecture reported decreased mind wandering (Wammes & Smilek, 2017).

Curiously however, the Wammes and Smilek study did not find evidence to support that mind wandering affected academic performance poorly, and it remains to be seen whether mind wandering actually inhibits learning. A comprehensive review of the costs and benefits of mind-wandering was conducted by Mooneyham and Schooler, and had specific interest in the context of learning and the question about whether mind wandering inhibits it (Mooneyham & Schooler, 2013). The review found that mind wandering negatively impacts reading and aptitude performance, attention, model building and working memory performance. This is a considerable span of outcomes, and though many of these were explored by the other theories, it lends considerable evidence for the role of mind wandering in the education process.

Perhaps the best evidence for this was the review's finding that mind wandering plays a positive role in autobiographical planning and creative problem solving. In a later study, this relationship was further corroborated, as researchers found potential benefits of mind wandering to creativity, specifically in the online learning and information technology context (Sullivan et al., 2015). Rather than simply being a negative phenomenon, mind wandering can have positive benefits. Recent research on this subject expands on these findings and suggest that mind wandering may not be detrimental to learning at all (Wammes, Seli, & Smilek, 2018). As Wammes et al. argue, the distinction between intentional and unintentional mind wandering may be to blame, as much of the extant research does not make this distinction. In their work, they find that students who experience mind wandering exhibit poor task performance, while unintentional mind wandering was actually associated with higher test scores in the long run.

Though mind wandering is a distinct observable phenomenon from the first person perspective, the considerable similarity among the observed learning impact and the other theories thus far explored should give us pause. The precise reasons why mind wandering negatively impacts learning remains unclear, in part because mind-wandering is not a concrete concept. Smallwood and Schooler shed light on this issue and asserts that mind wandering can be incorporated into models of executive attention and attentional control (Smallwood & Schooler, 2006). They observe that mind wandering is likely to occur when the primary task does not require executive control, and that the negative relationship between mind wandering and executive functioning supports such a relationship. They speculate that mind wandering involves redirecting executive attention away from the primary task to personal goal-oriented processing.

2.1.5 The Attention Lynchpin

This discussion about the relationship between attention and these four distinct areas of learning research leads us to the core subject of this dissertation. All four of the learning theories explored in this review hinge on a learner's ability to pay attention. The active processing assumption of the Cognitive Theory of Multimedia Learning requires that participants are attending to the lesson. Active learning techniques are designed to draw and maintain participant attention. The flow state is rooted in a deep state of attentiveness where a subject is intrinsically motivated to participate in an activity. Mind Wandering disrupts flow and is a state of reduced attention to the task. The concept of mind wandering is therefore fundamentally understood to be a function of our attentional processes. Participant attention is clearly fundamental to the successful attainment of learning outcomes and could play an essential role in the explanatory power of prominent e-learning theories. It thus follows that an inquiry into the role of attention in the success (or failure) of MOOCs would have significant implications for the design or implication of future MOOCs.

This should not be surprising to us. Many of the learning theories explored thus far come from a cognitivist paradigm, which hold that learning is fundamentally an association building process. Attention has been identified to play a critical role in this process, serving as an intermediary between stimuli and the formation of associations (Mitchell & Le Pelley, 2010). We thus have a string motivation for studying attention.

In order to pursue an inquiry into the relationship between MOOCs and attention however, we must first construct a functional understanding of the concept attention in the first place. As we shall see, this inquiry will reveal a gap in the literature which can only be filled by stretching across the subjects of cognitive neuroscience, information systems, and machine learning. We shall now explore the neurophysiology of attention before moving into how attention has been observed in Information Systems research and how new technologies can be constructed to measure it.

2.2 The Neurophysiology of Attention

Attention is an enduring subject of inquiry, particularly in Psychology and Neuroscience. Before its contemporary conceptualization, it was investigated by philosophers, particularly by early modern philosophers in the Rationalist and Empiricist traditions (Mole, 2017). Descartes in *Meditations*, for instance, alludes to the phenomena of attention and memory to justify how clear and distinct ideas cannot be doubted (Descartes, 1984). 17th century Empiricists such as Locke argued that all knowledge is rooted in observed phenomena, and that attention is one of the mechanisms that makes knowing possible (Locke, 1689). To this day attention continues to play a crucial role in philosophical discourse about the nature of consciousness and its role in the human mind.

Psychological conceptions of attention begin with William James, one of the pioneers of the field. James investigated the phenomenology of attention and identified it as a single function to focus on "one out of what seem several simultaneously possible objects or trains of thought" (James, 1890). Distinct from his concept of the *stream of consciousness*, James conceptualized attention is the mechanism by which particular experiences are singled out from other experiences. Extending beyond the philosophical empiricists' understanding, James also identified attention as a distinct process of the sensory organs, where such organs adjust to particular objects within the greater stream of consciousness. This view of attention benefits from being intuitive and simple, which may be one of the reasons why it is so enduring.

However, modern neuroscience has demonstrated that attention is significantly more complex than James' conception. Rather than a single mechanism, attention is a number of cognitive processes that work together to yield the attention phenomenon. Though there is some evidence to suggest that different organs do indeed have different attention mechanisms (eg. auditory, visual), there is significant evidence to suggest the existence of distinct brain networks that govern the sub processes of attention (Montemayor & Haladjian, 2015). Though there are different models of attention, we will focus on the the Posner attention networks model, a well-established and perhaps dominant model, which is grounded in extensive experimental evidence in visual and auditory attention (Posner & Rothbart, 2007; S. E. Petersen & Posner, 2012).

2.2.1 The Attention Networks Model

An attention network is a network of neurons that govern one of the functions of attention. We can learn about the existence of an attention network using neuroimaging, but we need not do so. The original attention networks model advanced by Posner was established by observing cognitive functions. This is what Posner did in his groundbreaking *Orienting of Attention* where he established the distinct orienting function (Posner, 1980). Many of the studies used to justify this conclusion used a cuing task designed to compare reaction times to stimuli. By comparing reactions of healthy and brain injured humans, Posner was able to establish the existence of a distinct neural mechanism for attentional orienting long before the development of fMRI.

These accounts of distinct attention networks eventually led Posner to propose a three function model, where the attention phenomenon is composed of three interdependent but distinct functions: alerting, orienting and executive control. *Alerting* describes the function of maintaining a high degree of sensitivity to stimuli, but is also distinct from general arousal. *Orienting* describes the process of aligning with the source of sensory signals. *Executive control* functions govern the resolution of conflict among stimuli, and which stimuli to focus on. Though there are other models, these three distinct functions continue to form the foundation of a number of ongoing research programs in attention and neuroimaging, as well as clinical work.

The alerting network was originally identified by observing sustained vigilance in behavioural studies and was eventually correlated with brain stem activity and networks in the right hemisphere (Posner & Rothbart, 2007). Though scholarship on the alerting functions has expanded from the original model, much of the research has corroborated alerting as a distinct network (S. E. Petersen & Posner, 2012). In most experiments and in real world scenarios, the alerting function is almost always observed in conjunction with orienting. This has led some scholars to question the independence of the networks (Fan et al., 2009). However, a comprehensive review on the role of norepinephrine and acetylcholine has been conducted for its role in ADHD maintenance. Norepinephrine has been observed influencing orienting functions, but not alerting (Beane & Marrocco, 2004). This lends considerable support for the notion of alerting as a distinct function of the brain, even if it interacts highly with orienting functions.

Orienting governs the fundamental mechanism of feature selection, an important and somewhat mysterious phenomenon which was the original topic investigated by Posner (Posner, 1980). In Posner's original conception of the network, orienting was observed in association with a distinct brain region in the pulvinar and superior colliculus (Posner & Petersen, 1990). Recent work on this subject has challenged the simplicity of the original conception, and suggests that orienting is much more complex. Corbetta and Shulman have reported orienting to be associated with a frontoparietal network and the dorsal system, where the frontoparietal network works as a sort of "circuit breaker" for the dorsal system, which generates the salient events (Corbetta & Shulman, 2002). Other research has found all types of orienting to be associated with frontal eye field activity in the frontal cortex (Thompson, Biscoe, & Sato, 2005). This suggested that visual attention may have a distinct premotor ocular component. However, recent research has called the distinction of unique visual orienting networks into question (Smith & Schenk, 2012).

Orienting is often further contrasted between overt orienting and covert orienting in addition to visual/non-visual. Overt orienting is the act of attending to a specific location by moving the eyes to that location. Covert orienting is the act of shifting attention without physically moving eyes (Posner, 1980). There is considerable evidence to suggest that these are distinct, if highly similar neural networks, or rely on a single network and distinct additional mechanisms (R. Klein, Kingstone, & Pontefract, 1992; Corbetta et al., 1998). More recent work has largely supported this view. For instance, some research has found that eye movements enhances object discrimination. This suggests that overt orienting follows either an enhanced or distinct neural process (Harrison, Mattingley, & Remington, 2013).

Of the three attention functions conceptualized by Posner (1990), executive control has been most subjected to change. Executive control was originally conceptualized to describe target detection, and in doing so explain the limited capacity of attention. Posner and Petersen found this function to be associated with connections between the medial frontal and anterior cingulate cortex (Posner & Petersen, 1990). These findings are intuitive, as these are also areas of the brain that are also correlated with decision making and reward. Recent understandings of executive control have expanded on this original conception considerably. Prominent theories suggest at least two separate executive control networks, as evidenced by studies which reveal distinct frontoparietal and cingulo-opercular networks (S. E. Petersen & Posner, 2012; Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008). Top-down regulation, which are brain mechanisms that govern other brain mechanisms, have been shown to play a role in the executive control process.

There are many other theories of executive control that have emerged since. One conception that is particularly relevant to this work is the conception of executive control as a function of working memory or as a component of the working memory network. The reasoning is that executive control has a capacity, so it would make sense if it was a function of working memory. Though some theories of executive control and working memory recognize it as many distinct smaller networks, each for different domains (i.e. visual, auditory) (Luck & Vogel, 2013) the dominant view seems to be a single underlying mechanism that unites them. Engle posits working memory as equivalent to executive attention as evidenced by the correlation between working memory and attentional control tasks (Engle, 2002). Other evidence for this relationship is the strong correlation between working memory capacity and executive functioning constructs, suggesting a common underlying executive attention component (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010).

If this is true, the distinction between the three attention networks has implications for learning experiments and similar information systems problems. Working memory capacity, which is the underlying mechanism behind cognitive load, can be conceptualized as a significant part of a student's ability to attend to a stimulus. The dominant multimedia and active learning theories were thus correct to account for this, and may account for executive attention well. However, this is only one part of attention. The learning theories did not account for alerting and orienting functions, and these may play a large role in the successful design of learning multimedia.

2.2.2 Measuring Attention Networks Using Cognitive Tests

As mentioned previously, attention networks have been primarily measured by using cognitive tests that observe responses, such as reaction time, to stimuli. The Posner cuing task, for instance, was developed to measure and demonstrate the existence of a unique orienting network (Posner, 1980). More recently, the Attention Network Test (ANT) created by Fan et al. has been the most prominent example of a method to test the distinct attention networks (Fan, McCandliss, Sommer, Raz, & Posner, 2002; Fan, McCandliss, Fossella, Flombaum, & Posner, 2005). The ANT uses a combination of flanker tests, which are designed to measure response inhibition, as well as cuing tasks demonstrated to measure orienting. In the ANT, participants are asked to respond as quickly and accurately as possible to the tests, and these measures are used to measure the efficiency of the respective systems. The attention networks are differently engaged by uninformative (alerting) or informative (orienting) cues, as well as by flanking arrows that are either congruent or incongruent with the target stimulus (executive control).

There are some limitations with the ANT as developed by Fan (2005). The first is that the alerting and orienting networks are both observed though the cue condition, meaning that we cannot identify potential interaction between the two networks. Further, the nature of the task does not account for potential exogenous and endogenous components of orienting, given that the attention is fixated at the center of the screen (R. M. Klein, 2004; Ishigami & Klein, 2009). Second, though the ANT has been consistently demonstrated to be a reliable attention measure (Ishigami & Klein, 2011; Ishigami et al., 2016), there are components of attention, particularly executive attention, that it does not take into account. In response, researchers have developed alternatives to the original ANT that account for some of these criticisms (Callejas, Lupianez, Funes, & Tudela, 2005) or have developed expanded tests. The Dalhousie Computerized Attention Battery (DalCAB) is an example of such an attention battery, which uses eight reaction time tests, with additional measures such as vigilance (Jones et al., 2016). Moving forward, ANT and related cognitive tests can provide useful measures of attention that can be correlated with learning contexts.

2.2.3 Investigating Attention Networks using Neuroimaging

Much of the work on attention networks is validated through neuroimaging such as functional magnetic resonance imaging (fMRI), but also using positron emission tomography (PET), electroencephalography (EEG) and magnetoencephalography (MEG). Perhaps the most notable neuroimaging studies on attention networks was completed by Posner and Petersen before the advent of fMRI. (Posner, 1987; S. Petersen, Fox, Miezin, & Raichle, 1988; Posner & Petersen, 1990). By interpreting the results of animal studies, behavioural studies and on human participants with disorders caused by brain injury, researchers advanced the three networks model previously described. Later, when Posner revisited this early work, he reflected that they were surprised evidence for this model gained evidence and additional support over time, especially when fMRI became mainstream (Posner & Rothbart, 2007).

After the discovery of fMRI however, much of the research investigating attention networks came from fMRI brain imaging. As previously described, work by Fan et al. (2005) on the attention networks test has given the Posner and Petersen model additional support. Alerting has been found to be associated with thalamic and anterior activation. Orienting is found at parietal sites, as well as near ocular regions of the brain. Executive attention is observed in the anterior cingulate. Given the benefits of source localization and the identification of attention networks, there are clear benefits to using fMRI in attention research. However, fMRI still has some limitations, especially the limited number of applications that can be tested in an fMRI environment. In IS research in particular, electroencephalography (EEG) (and consequently MEG) correlates may be more useful because of their potential in ecologically-valid IS contexts (Riedl & Léger, 2016). It may also be able to give more insight into changes in attention over time given the poor temporal resolution of fMRI.

The P1-N1-P2 complex complex is a mandatory event-related potential (ERP) response triggered by early attention control mechanisms. This complex is sensitive to both visual and auditory stimuli, and is observed in the occipital regions of the brain (Hillyard, Hink, Schwent, & Picton, 1973; Hillyard, Vogel, & Luck, 1998). When a stimulus is detected by the auditory or visual system, this pattern of electrical potentials can be observed at 100-220ms. Attended auditory stimuli can be observed

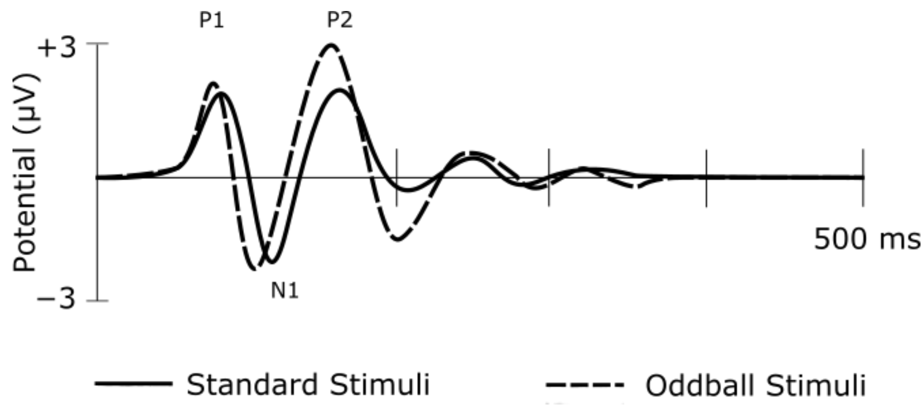


Figure 2.1: A hypothetical P1-N1-P2 complex waveform

having higher electrical amplitudes from this response (Teder-Sälejärvi, Münte, Sperlich, & Hillyard, 1999). This response can also be observed having higher electrical amplitudes to so-called oddball paradigm stimuli, which are stimuli that are different from those repeated throughout the duration of an experiment (Luck, 2014). Early negative electrical potential responses from this complex have been found to be associated with alerting, and have been observed during the Attention Network Test (Posner & Rothbart, 2007). The P1-N1-P2 complex is thus a useful neurophysiological response that can be used to observe alerting and orienting networks in the context of human-computer interactions.

A second EEG component that is common in attention research is the P3 component component. Where the P1-N1-P2 complex is triggered by alerting and orienting functions at the onset of all detected stimuli, the P3 component is associated with later attention. The P3 response occurs immediately following the P1-N1-P2 response, typically between 250-500 ms, and has been demonstrated to be driven by the activation of orienting networks (in the case of the P3a subcomponents) or of executive attention and updating in working memory (in the case of the P3b subcomponent) (Polich, 2007). Research with the ANT has demonstrated that the P3 is also evoked during cuing tasks, and is observed in association with attention capacity (Petley et al., 2018). The study of the P3 response can thus also be a useful neurophysiological indicator to observe executive functions.

A final measure that can be detected with electroencephalography (EEG) are the patterns of oscillatory activity. Neural oscillations are the thought to be the result of the central nervous system, and have been noted for their usefulness in measuring constructs that are relevant to IS research, such as vigilance or relaxation (Riedl & Léger, 2016). In attention research, oscillatory activity activity at the beta (13-30 Hz) and gamma (30-100 Hz) bands has been found associated with executive attention, as well as working and long-term memory activation (Tallon-Baudry, Kreiter, & Bertrand, 1999; Jensen, Kaiser, & Lachaux, 2007). Lower frequency alpha band activity (7-12 Hz) have also been observed in conjunction with orienting and attentional suppression (Klimesch, 2012), suggesting that the activation of attention mechanisms may be associated with higher band activity.

Oscillatory measures have also been observed in the context of mind wandering. Braboszcz and Delorme (2011) observed increases the band activity correlates between mind wandering and on-task states during a meditative task. They observed increased oscillatory activity at the delta and theta bands and decreased activity at the alpha and beta bands when participants reported mind wandering. However, other studies have found different associations, such as a relationship between increased alpha activity and mind wandering during a driving task (Baldwin et al., 2017), as well as increased somatosensory alpha rhythms when successfully avoiding mind wandering during a meditation task (Brandmeyer & Delorme, 2018). This suggests that the presence of mind wandering may generate different frequency band activity depending on the task that participants have been given. Activation of attentional mechanisms, and consequently the alpha activity associated with it, may be useful correlates of mind wandering or on-task states depending on the experimental task.

2.3 Bridging the Gap Between Neuroscience and Information Systems

Thus far we have explored learning theories and the neurophysiology of attention. However, as mentioned in the introduction, this dissertation is fundamentally about Information Systems (IS) and merely draws on this knowledge to achieve IS research goals. The subject of Information Systems is an applied discipline, focusing on sociotechnical systems compromising on the interaction between humans and information technology. In IS scholarship, it can be easy to be distracted by the technology or reference discipline, and to lose sight over the fact that IS is fundamentally a social science focused on technology use (Recker, 2012). As with all social sciences, IS scholars must cope with the imprecision and ambiguity implicit with social science endeavours. This has motivated IS researchers to create methods that primarily focus on predicting human behaviour rather than explaining its natural causes. IS scholars are often content to observe latent variables rather than the natural forces that are the foundation of those variables in the first place. Though this is a generalization of an entire discipline that by no means applies to all IS research, it most certainly applies to the portion of IS research conducted on the topic of attention. Though there are many potential applications for improved attention measures, perhaps the most promising is the improvement of existing IS constructs. We shall briefly explore some of these constructs and suggest how the discussion thus far can inform their improvement. We will then discuss the specific constructs explored in this dissertation.

2.3.1 Attention-Related Information Systems Constructs

There are a number of IS contexts where the study of attention is relevant. Perhaps most notably, attention has been interesting to IS researchers working in the context of e-commerce, where it has also been studied as a significant factor in visual search. The first study where this is found observes an attentional capacity construct (Hong, Thong, & Tam, 2004). Drawing on the work of Khaneman, the study's attentional capacity construct was designed to account for central capacity theory, which describes how when mental resources are spent suppressing a stimulus, such as the interference of a flash animation, less attention will be available to process non-flashed items, consequently limiting participantś ability to recall the information they were tasked with (Kahneman, 1973). Similar to extraneous cognitive load as described by Sweller and the CTML described by Mayer (Sweller, 1994; Mayer, 2009), Kahnemanś theory envisions attention as a limited capacity. The attentional capacity construct described by Hong et al. (2004) uses a combination of ex post questionnaire measures related to cognitive absorption and concentration, as well as success rates at an information recall task to account for limitations in attentional capacity. In hindsight, the phenomena observed by Hone et al. (2004) could also be explained by cognitive load theory, and not necessarily by drawing from attentional processes such as orienting or executive control as described thus far. Later studies in this subject have taken similar approaches to attention, drawing from Kahneman or related literature. For instance, attentional capacity has been used to explain variances in e-commerce task performance and measure using total numbers of clicks (Tam & Ho, 2006) or by the results of online auctions (Tan, Yi, & Chan, 2008).

Another topic of interest related to e-commerce has been the subject of online wait times. Lee et. al (2012) likewise drew from resource allocation theories of attention, but also "competition for attention" theories from consumer research (Lee, Chen, & Ilie, 2012; Janiszewski, 1998). These theories hold that focal and non-focal objects compete for attention in a visual field, and that the way stimuli are presented will impact attentions' limited capacity. The result was an attentional capacity construct that could account for the impact of online wait times on bounce rates. Curiously, Lee et al. note that only a handful of IS researchers have successfully applied this theory to find a significant effect (Lee et al., 2012). The following year, Hong et al. would call for a new way for e-commerce researchers measure attention in the context of wait times, focusing on the subjective experience of waiting rather than the objective wait times themselves (Hong, Hess, & Hardin, 2013). They have inconsistent findings on the impact of objective wait times, but also found that the amount of visual content provided during the waits had an impact on user perceptions.

Attention has also been studied outside of e-commerce contexts. Dabbish and Kraut conducted human-computer interaction experiments to determine optimal awareness displays for summarizing team members' workloads (Dabbish & Kraut, 2008). In that study, the authors explored the role of attentional demand, a construct described by the authors to account for how increasing the number of visual elements and movement makes visual stimuli more distracting. Grounded in perceptual psychology (Pashler, Johnston, & Ruthruff, 2001; Wickens, Gordon, Liu, & Lee, 1998), the attentional demand construct was used to account for how differences in the amount of information presented to the user impacts visual task performance and was measured using the proportions of eye fixations focused on distracting display features versus relevant features. Unlike the aforementioned attentional capacity construct, task performance was explained by eye fixations, which may be related to where attention is oriented and executive attention fixated to particular stimuli. In a similar study, Potter and Balthazard investigated the role of attention in electronic brainstorming software (Potter & Balthazard, 2004). They demonstrated investigated the impacts of different types of cues on the successful completion of brainstorming tasks. Similarly, when executive attention is distracted and drawn away from the task, brainstorming productivity was hindered.

The primary method used in these examples were task success measures, which can be observed in each of the examples cited above. Additionally, psychometric questionnaire measures were employed in two studies (Hong et al., 2013; Lee et al., 2012) while comprehension measures were also employed (Tam & Ho, 2006). It is clear that the primary methods used to observe attention-related constructs have been behavioural tasks and ex post measures, which are also characteristic of measures typically employed in IS research as a whole. Alternative approaches to observing these constructs could involve examining them using neuroimaging or electroencephalography (EEG), specifically by examining the neurophysiological correlates of attention in an IS setting. However, it is not immediately clear why we would do this, especially considering that most IS goals can be accomplished using traditional measures. We shall now explore why it is desirable to use neurophysiological measures for IS research in the first place.

2.3.2 The NeuroIS Research Agenda

As mentioned earlier in this chapter, this dissertation is firmly part of the "Neuro-Information Systems" (stylized NeuroIS) movement. NeuroIS is a new subfield within IS research, with its origins dating back to a 2008 article by Dimoka and Davis which explored the neural mechanisms behind the seminal *Technology Acceptance Model* (Dimoka & Davis, 2008). That same year, the first NeuroIS retreat was held in Gmunden, Austria, which aimed to usher in a movement to conduct new IS research using methods from cognitive and affective neuroscience. Subsequent work at the Association for Information Systems Special Interest Group on Human Computer Interaction (SIGCHI) (Riedl, Randolph, vom Brocke, Léger, & Dimoka, 2010), *Information Systems Research*, (Dimoka, Pavlou, & Davis, 2011) and *MIS Quarterly* (Dimoka et al., 2012) explored the methodological and technological foundations of such an endeavour. These early works largely focused on establishing the legitimacy of NeuroIS as a subfield by emphasizing how techniques can complement existing IS research tools, capturing the antecedents of IS constructs, or by enhancing existing IS theories.

Many of these assumptions continue to inform NeuroIS research today, though the motivations have expanded. Some researchers such as Riedl and Léger have posited that neuroscience does not merely need to play a complementary role, but can also directly inform the design of IT artifacts, particularly in the context of graphical user interfaces (GUIs) (Riedl & Léger, 2016; Brocke et al., 2013; Riedl, Davis, Banker, & Kenning, 2017). Hypotheses about GUIs could be tested using neuroimaging in information technology settings, and directly informs the design of socio-technical systems. Given a physiological, rather than subjective justification, engineers could have better justification for applying such findings. Furthermore, though neuroscience can certainly complement existing subjective measures by triangulating them with traditional IS observations, there are some measures that cannot be easily observed subjectively. Some measures, such as trust or attention, may be biased by subjective report. Others may be impossible to measure using questionnaires because interruption may invalidate the phenomenon being measured. This motivation informs one of the core arguments of this thesis, which is that mind wandering is one such measure, which cannot easily be measured subjectively and interruption may invalidate the phenomenon being measured.

The history of NeuroIS as of 2017 is meticulously described by Riedl et al. in an article presented at the International Conference for Information Systems, which we will not expand on further (Riedl, Fischer, & Léger, 2017). At the subsequent 2018 NeuroIS conference, Fischer et al. presented the results of a survey study on the current state of NeuroIS research (Fischer, Davis, & Riedl, 2019). They found that five have dominated past NeuroIS research: stress, trust, cognitive load, emotion and attention. That survey also queried participants about future research focus, and found that that the overwhelming majority of respondents were interested in either emotion or cognitive load. The survey also found a division between American and European researchers. European researchers were either conducting or aspired to conduct future research in emotion or stress, while American researchers were interested primarily in cognitive load and design science. Curiously, American researchers were also more likely to identify Attention as a topic of interest than European researchers. We can speculate that this is because attention informs cognitive load and design science, which are the primary applications of research among most of the North American NeuroIS groups.

2.3.3 Attention Constructs and Neurophysiology

Though we are no aware of any extant work in the IS literature that leverages the neuroscience of alerting or orienting, we are by no means the first to investigate attention-related neurophysiology in the IS context. Eye tracking is commonly employed neurotechnology to IS attention research (Riedl, Fischer, & Léger, 2017). In a seminal paper, eye tracking was used to investigate user interest and attention to e-commerce images (Cyr, Head, Larios, & Pan, 2009). Ultimately, the authors demonstrated that eye tracking could be effectively used to identify interesting images, but did not specifically identify that attentional processes were being observed. Other researchers have used gaze fixation to measure differences in cognitive load in reading contexts (Liu, Lai, & Chuang, 2011), to measure attention to specific web features (Djamasbi, Siegel, & Tullis, 2012), or to measure attention during information search (Cole, Hendahewa, Belkin, & Shah, 2015). These examples are relevant examples of overt attention measurement, but may not be appropriate in all contexts. In the context of MOOCs, users can fixate in any number of locations and still be in a generally attentive state.

Alternative methods have been employed to measure attention covertly in more similar contexts. As already discussed, the neurophysiological correlates of cognitive absorption have been observed and reported by Léger et al. (Léger et al., 2010; Léger, Davis, et al., 2014), who studied the role of cognitive absorption in an enactive gaming context. In this study, EDA as well as EEG oscillatory activity were employed to observe changes in users' cognitive absorption states. The authors justified the use of electrodermal acivity (EDA) by relating it to the literature on arousal, which has been closely associated with the alerting network (Posner & Petersen, 1990). Additionally, mid-range frequency activity at the alpha (7-12 Hz) and beta (13-30 Hz) bands were utilized to observe the cognitive absorption state. Alpha activity, which is commonly associated with alerting attention, was observed as being positively associated with the latent cognitive absorption construct. EEG beta activity, which is commonly associated with cognitive activity, was observed in negative association.

In a paper published the same year, researchers addressed the neurophysiology of attention directly by investigating the P3 event related potential and its correlation with eye-fixation related potential (EFRP)s (Léger, Sénecal, et al., 2014). The emphasis of this study was to establish measures of attention that are appropriate in ecologically-valid IS contexts, and ultimately concluded by establishing the validity of EFRP in IS research contexts. Though eye-tracking methods reflect the state of the art in orienting research (Geva, Zivan, Warsha, & Olchik, 2013), the fact that these measures were justified though association with the P3 component event-related potential (ERP) suggests that they are measures of *overt* orienting, and not the *covert* orienting described by Posner et al. (Posner & Petersen, 1990). The EFRP thus described is thus very useful in attention research, but mainly in contexts where eye fixation is the relevant measure. Covert orienting, by contrast, remains a potential topic of interest of IS research, especially research that does not involve a visual component. Such questions may benefit by leveraging covert orienting measures, which we will soon explore.

In addition to attention to the work from colleagues at HEC Montréal, there has been considerable recent interest in covert attention measures and attention-related constructs among the attendees of the NeuroIS retreat. Shamy and Hassanein are currently exploring effects of attention-related measures on age (El Shamy & Hassanein, 2018) and are using EEG measures to investigate whether attentional overload impacts creativity (Calic, El Shamy, Hassanein, & Watter, 2018). Eye tracking is currently being used to investigate attention and behaviour with security warnings (Vance, Jenkins, Anderson, Kirwan, & Bjornn, 2019), for detecting mind wandering during web tasks (Gwizdka, 2019) and during anomaly detection during software use (Boutin, Léger, Davis, Hevner, & Labonté-LeMoyne, 2019). We can safely state that there is not only precedent for exploring the neurophysiology of attention in IS contexts, but also that there is considerable present interest in this subject among IS scholars.

When it comes to the specific measures of mind wandering, we have already seen that mind wandering can be somewhat effectively measured using ex post questionnaires. However, these methods offer limited insights into their cognitive and neurophysiological correlates. In addition, they do not offer us much insight into the temporal nature on the phenomenon, and suffers from many of the drawbacks already discussed in the literature on flow. One method for overcoming this limitation is to use experience sampling, which is a method of employing a simple yes/no measure in order to determine the occurrence of mind wandering (Schooler, 2004; Smallwood & Schooler, 2006). Using this method we can measure the presence of the phenomenon in real time. However, it suffers from the disadvantage of disrupting the very phenomenon we are trying to measure.

An alternative is to use neuroimaging. Hillyard et al. (1998) describe an EEG auditory oddball paradigm paradigm which triggers the P1-N1-P2 complex, the series of three electrical peaks triggered by early attention control mechanisms described earlier (Hillyard & Anllo-Vento, 1998). This complex is an indicator of the switch of general selective attention towards a stimulus. Consequently, explicitly attended stimuli elicit larger amplitudes relative to unattended stimuli. The mechanism has been well-tested and has been associated with other correlates of attention using fMRI (Hillyard et al., 1998).

Braboszcz and Delorme (2011) described an experiment to measure mind wandering using both P1-N1-P2 complex oddballs and oscillatory activity. They asked participants to push a button when they experienced mind wandering and found heightened P2 amplitude, as well as greater delta and theta band power and lower alpha activity when participants experienced mind wandering versus when they were on-task (Braboszcz & Delorme, 2011). Such measures may be applied to the e-learning context, and could eventually yield a significant measure of mind wandering that can be employed without disrupting the user.

One of the potential weaknesses of the experience sampling technique described by Braboszcz and Delorme is that they are binary, and hinge on an individual's ability to detect when it occurs. Though they employed a counting task to give an objective measure of when mind wandering occurs, studies that similarly employ experience sampling without such a task run the risk of failing to measure the precise point when the mind wandering phenomena occurs. An alternative approach is to deliver "thought probes", which are questionnaires that occur at fixed times. Wammes and Smilek employed this technique when investigating the effect of mind wandering in the classroom environment (Wammes & Smilek, 2017). In addition to overcoming the timing issue, this method allows researchers to measure the degree of experienced mind wandering by asking participants to complete a Likert scale questionnaire. Such techniques could be used to overcome the limitation of subjective mind wandering measures and by measuring the degree of experienced mind wandering. When combined with neurophysiological data, we could use observe correlates between the degree of reported mind wandering and the degree of measured P2 amplitude.

Other research has observed variances in the P3 component in relation to mind wandering. Smallwood et al. described an experiment where participants were asked to attend to visual stimuli while reporting experienced mind wandering (Smallwood et al., 2008). Visual stimuli triggered the P3 attention response, and exhibited lower amplitudes when in states of reported mind wandering. The authors inferred that this phenomenon represents the direction of overt attention away from the task and towards self-directed thoughts. An alternative approach to observe responses to auditory stimuli was recently taken in the context of simulated driving (Baldwin et al., 2017). Participants in states of reported mind wandering exhibited lower P3 responses to auditory cues. In this later study, mind wandering was also found to have a negative impact on driving performance which highlights the importance of understanding the mind wandering phenomenon.

Finally, combining neurophysiological measures with subjective measures allows researchers to generate data labels, which can be used to create a machine learning algorithm which detects attention-related states in real-time. This also raises the possibility of an attention-adaptive brain-computer intferface (BCI). BCIs have been proposed and employed in Information Systems research before though they are not yet widely adopted (Randolph et al., 2015). In the case of mind wandering, self-reported measures could be used to create labels for machine learning classification. If such machine learning classifiers could be demonstrated to accurately classify event-related potentials, the result would be a measure that is able to detect mind wandering passively in real-time. To the best of our knowledge, this has not been done before, and would represent a novel contribution to both Information Systems and to the literature on mind wandering as a whole. We will explore this concept in detail, but must first explain how electroencephalography (EEG) data is used to conduct such research in the first place.

2.4 Techniques for Analyzing EEG Data

We have explored learning theories and the role that attention plays in them. We have also discussed the neurophysiology of attention, have established that EEG measures can be used to effectively measure covert attentional orienting, and that these can also be used to measure mind wandering in e-learning contexts such as online lectures. The remaining task is to describe how data retrieved from EEG can be used to accurately measure neurophysiological states, and how it can do so in real time. This conversation will discuss the details of two technical domains that seem to be unrelated at first glance. The first is the art of EEG analysis, and specifically the event-related potential (ERP) technique. The second is machine learning, which can be used to process and interpret electroencephalography (EEG) data. The two domains are inter-related to the extent that machine learning informs ERP detection science and data processing, but also makes it possible to create real-time measures which are capable of interpreting changes in attention states. We shall explore each of these three concepts in turn, with attention to the specific applications that inform the analysis in this dissertation.

2.4.1 Conventional EEG Analysis

Electroencephalography is a relatively old technology, at least from the perspective of neuroscience. The systematic study of EEG can be traced back to 1929, when Hans Berger coined the term "electroencephalogram" to refer to refer to devices which record electrical potential on the human scalp (Haas, 2003). In the early days of electroencephalography (EEG), the tool was mostly used to observe patterns of oscillations, which are readily visible to the naked eye. Well-defined neural oscillations, especially those at the theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz) bands, are an obvious visual feature in EEG data and are present in most sustained EEG observations. Research into the relationships between event related potentials

(ERPs) and cognition did not begin earnestly until the 1960's however, as the instruments remained too crude to detect the comparably smaller event-related potentials. Starting with the discovery of contingent negative variation, an ERP triggered by subjects' anticipation of events, there was excitement in the ERP technique's potential (Walter, Cooper, Aldridge, McCallum, & Winter, 1964). The following year, the attention-related P3 component was discovered (Sutton, Braren, Zubin, & John, 1965), triggering a flood of ERP research (Luck, 2014).

What are ERPs? ERPs are the result of postsynaptic potentials triggered by neurons in response to stimuli. When a large number of neurons trigger simultaneously, the electrical potential aggregates and is conducted through the brain and scalp. Given that the ERP is the result of electrical activity, the result is a near-instantaneous measure of neurotransmission. Unfortunately however, this is the source of both ERP's greatest advantage, as well as its greatest disadvantage. The disadvantage of measuring such brain-sourced electrical activity is that only a tiny fraction of neural combinations ultimately result in a sizable enough dipole that is not equalled out by a similar process and terminates at the scalp. The greatest advantage is, given its electrical origin, EEG generally offers excellent temporal resolution when compared to fMRI or PET (Luck, 2014).

Neural oscillations, by contrast, are likewise the result of postsynaptic potentials, but at a much more general level. Where ERPs are triggered by specific events at fixed periods of time, oscillatory activity is the result of ongoing brain activity, which aggregates to create patterns of variance, often referred to as *eeg waves*. As mentioned earlier, different electroencephalography (EEG) wave frequencies have been associated with constructs such as beta and working memory (Roberts, Hsieh, & Ranganath, 2013), or even alpha and mind wandering (Braboszcz & Delorme, 2011). We should be cautious when making these associations however, as frequency activity reflects general, as opposed to specific neural processes. The trigger of an ERP is often well-defined, while EEG wave patterns may the result of completely unrelated processes such as motor activity or subject fatigue.

Raw EEG data consists of one time series recording for each applied EEG electrode, sampled at a given rate over a sustained experiment. EEG recording hardware can be configured to facilitate real time recording of events and brain data, while

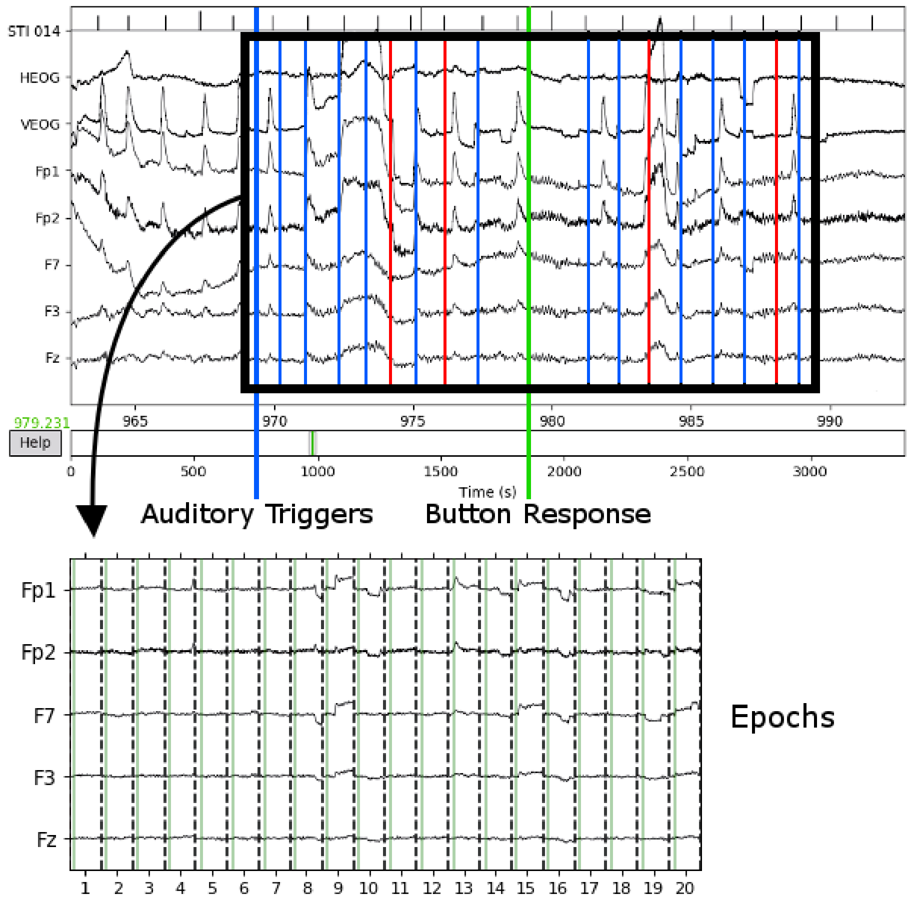


Figure 2.2: Raw EEG Data is Transformed into Epochs

limiting outside electrical noise. When it comes to data processing, given that we are primarily interested in short segments of data around these recorded events (as with ERP research) or in frequency patterns (as in oscillatory research). At the basic level, raw EEG data is typically divided into epochs; segments of time that are of interest to the research question. In traditional ERP research, this usually involves extracting 700-1200 ms intervals around the physically recorded event triggers, which represent stimulus delivery (Luck, 2014). EEG electrodes detect electrical activity at the scalp, much of which is generated by muscle activity, rather than brain activity. Blinks and jaw clenches, for instance, are highly visible artifacts detected by EEG sensors. One approach is to manually remove epochs that feature such artifacts.

This does not work for all artifacts however. For instance, the noise created by electrical devices at the 60 Hz frequency cannot be removed this way. An alternative approach is to apply a spatial filter the data. A common method to filter data of noise is to apply a high and low bandpass filters, which filter signals below 0.1 Hz and above 50 Hz. Data can be further filtered using unsupervised machine learning techniques such as independent component analysis (ICA). Abstractly, ICA interprets a given vector of observations $x = (x_1, ..., x_m)^T$ using a static linear transformation W into s independent components such that $s = Wx$. ICA then identifies the components as measured by a function of independence $F(s_1, ..., s_n)$. Practically, the components that most account for the most variance will usually be muscle or other noise artifacts, which can be rejected from the data. In some cases, supervised machine learning filters such as xDAWN may be applied, which uses the epoch labels to maximize the signal to signal plus noise ratio of an evoked response (Rivet, Souloumiac, Attina, & Gibert, 2009). Common electroencephalography (EEG) analysis tools such as EEGlab are designed to make it easy to apply these filters in the EEG context (Delorme & Makeig, 2004).

When quality data is filtered, the result will be a number of epochs for each observed condition that trend toward an event-related potential (ERP) waveform or actual frequency patterns as illustrated in Figure 2.2. Standard ERP analysis involves identifying the average waveform by averaging the results of a series of trials within a subject. The number of trials required in order to accurately calculate the ERP waveform depends on the type of waveform being observed, ranging from anywhere between 1-50 for the P3 component to 100-500 for the P1 component (Luck, 2014). To compound the complexity, there is also significant variance between subjects that should be accounted for. To account for this, ERP experts often recommend calculating a grand average, which consists of averages between multiple subjects (Delorme, Miyakoshi, Jung, & Makeig, 2015). By doing this, we can be more confident that the waveform accounts for variance between subjects, as well as between conditions. Tools such as MNE Python facilitate the calculation and visualization of ERP waveforms in the Python programming environment, while also maintaining a data structure that is favorable to leveraging other Python statistical and machine learning libraries (Gramfort et al., 2013, 2014).

When the waveforms are computed, they can be compared statistically. ERP waveforms are often defined by their amplitudes and latencies. In the early days

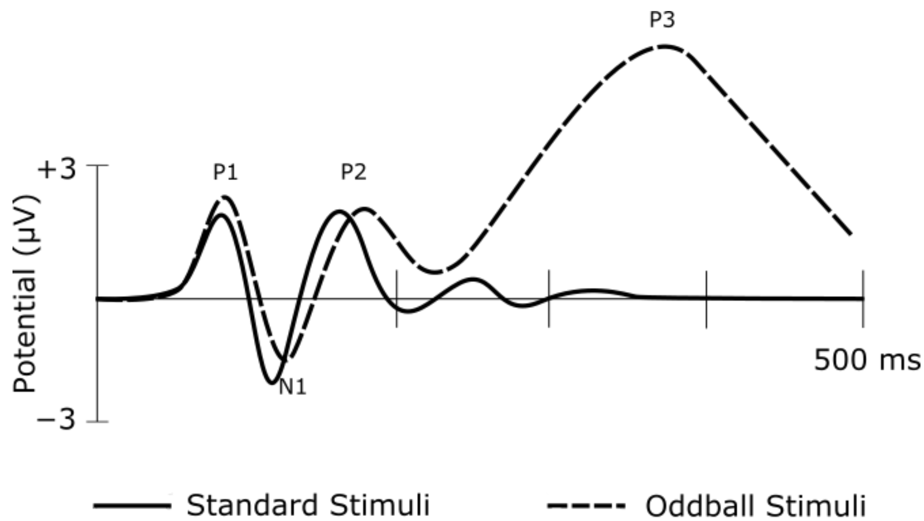


Figure 2.3: Hypothetical P3 oddball paradigm waveforms

of ERP analysis, specific peaks of a waveform were compared by calculating a defined measurement time window and comparing the peaks of the waveforms between conditions (Dunchin & Heffley, 1978). Though this technique is still used, there are arguments against it, and it has been shown that mean amplitude is usually a better alternative (Luck, 2014). Regardless, both peaks and amplitudes of the calculated waveforms can be analyzed, and compared statistically using a simple t-test. They can also be used as features in group-level analysis, and compared to survey instrument results using linear regression or ANOVA analysis. Another procedure is to use linear mixed effects models (LME), as these models are able to account for supposedly random effects, such as variance between subjects (Davidson, 2009). This helps establish the validity of results with a large number of trials (such as in ERP analysis) but also have a small number of subjects. It does this by accounting for inter-subject variance.

In addition to time domain ERP analysis, we can also conduct frequency analysis around the events being investigated. As mentioned earlier, electroencephalography (EEG) oscillations happen at different frequencies, and these are commonly broken down into frequency bands (i.e. "alpha" at 8-13 Hz, "beta" at 13-30 Hz). Using the Fourier transformation, we can transform a time-series signal into the frequency domain spectrum, which can be used to detect these patterns. We can formally express the Fourier transformation from time into the frequency domain as follows, where the independent variable x represents time and k represents frequency (in our case, Hertz):

$$f(x) = \int_{-\infty}^{\infty} F(k)e^{2\pi ikx}dk$$

Heightened activity at particular frequencies are often analyzed in relation to group features, and can be a meaningful predictor of phenomena. For example, Léger et al. (2014) investigated the relationship between activity at the alpha and beta band frequencies to determine a cognitive absorption measure, while Braboszcz and Delorme investigated time-frequency differences between mind wandering and on task states (Léger, Davis, et al., 2014; Braboszcz & Delorme, 2011). Likewise, by investigating time-frequency segments, we may find they are significant predictors of the phenomena being investigated.

2.4.2 Machine Learning for Real-Time Signal Detection

So far, our conversation has been about computing the results of EEG and making inferences about those computations. These calculations are conducted offline, and primarily employ statistics to do the job. An alternative approach is to create measures that interpret EEG data in real-time; a process often referred to as decoding (Haxby, Connolly, & Guntupalli, 2014). Decoded signals can theoretically be used to translate brain activity patterns into commands in an interactive system. This is what we mean by brain-computer intferface (BCI). As already discussed however, raw EEG data is complex and noisy, and it would be very difficult to explicitly create computer programs that interpret such signals. This is why BCIs largely employ machine learning to interpret the data. We have already discussed some of the applications of machine learning without having formally discussed what machine learning is. We will now do this, with special attention to the specifics that inform EEG and BCIs.

Formally, machine learning can be defined as "the field of study that gives computers the ability to learn without being explicitly programmed". This definition is normally attributed to Arthur Samuel, one of the field's pioneers, though it is unclear whether he actually said it. In popular culture, we often associate machine learning with intelligent robots, or machines with complex, human-like digital brains with general intelligences that exceed our own. Even generally educated readers are often familiar with recent advancements in deep learning, which often leverage neuron-like functions and vast amounts of data, feeding this popular conception. However, the concept of machine learning is much simpler. For instance, machine learning could involve the application of simple polynomial procedures or heuristics to infer a series of rules in a game of checkers (Samuel, 1959). What makes it different from regular programming is the way that rules adapts to the structure to the given data.

Machine learning algorithms are often broken into two categories: *supervised* and *unsupervised* learning (Flach, 2012). Supervised learning involves learning from data for which one has labels—indicators of the ground truth. For example, having labels of "spam" and "ham" emails would help supervised learning algorithms identify which emails are useful to users. Unsupervised learning, by contrast, does not require labels and tend to work very differently from supervised learning. These techniques often involve clustering, which is following a rule for grouping data. Other topics in machine learning include issues of how we prepare or structure data, and evaluate performance of algorithms. We will not discuss the broad details of this field, but will discuss two specific issues within machine learning: *data dimensionality reduction techniques* and *general classification techniques*, which are specifically relevant to real-time detection and BCI tasks.

EEG data is highly dimensional, especially considering the comparatively small amount of it typically analyzed in ERP research. For a standard 32 channel EEG system, there are 32 channels recording at 512 Hz; one data point for every 2 milliseconds. As mentioned earlier, though there are many methods for representing EEG signals, the two most common are time-series and frequency band power (Bashashati, Fatourechi, Ward, & Birch, 2007). For time-series analysis, we may therefore record hundreds of samples for each observed condition, but have thousands of features for each sample, when we consider the range of time being observed. Depending on how the time-series features are segmented, there are potentially dozens of features for each epoch.

The typical solution to the problem of dimensionality is to add a feature selection step that reduces the number of features in the data. Using a feature selection method to either extract the most relevant features or compress the features prevents overtraining effects and reduces the number of features that classifiers need to process. Lotte et al. (2018) identify three common feature selection approaches that are commonly used in BCIs: filter, wrapper and embedded approaches (Lotte et al., 2018). Filter approaches rely on measures of relationship between each feature, and transform the data correlating to those measures. Wrapper and embedded approaches instead use a classifier to obtain a subset of features, and either select a subset of features (wrapper) or integrate the features selected into a unique process (embedded). A famous example of embedded feature selection is to apply stepwise linear discriminant analysis (Krusienski et al., 2006). This technique attempts to explain differences in classes by identifying linear combinations of data features that best discriminate between the classes. It has been noted for its improvement of time domain BCIs.

Following dimension reduction, data can be classified. In its simplest form, classification involves employing algorithms that interpret data to discriminate between two classes, though it is by no means restricted to this binary form. The classification process involves two phases: a *training* phase and a *test* phase. During the training phase, classifiers build models by interpreting the data to discriminate between the assigned labels. During the test phase, model performance is tested on related data that was not used in the training phase.

The success of a classification model is often evaluated one of three ways: *accuracy*, *precision* or *recall*. Accuracy, being the most common measure, can be expressed in relation to test set Te as follows (Flach, 2012):

$$acc = \frac{1}{|Te|} \sum_{x \in Te} I[\hat{c}(x) = c(x)]$$

In this expression, I is either true or false, and is determined by whether the estimated class ĉ(x) is equivalent to the true class label c(x). Alternatively, we could express accuracy as a function of correct classification of true positives and true negatives in relation to all classified states (including false positives and false negatives). It can be defined this was as

$$acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy is often used to denote the estimated probability that a classifier will correctly classify instances of data. Precision, by contrast, measures the proportion of true positives among predicted positives. Precision is often employed in contexts where it is most important that the predicted values are correct. Alternatively, recall measures the total proportion of true positives in relation to false negatives. Recall measures are employed in contexts where it critical that all true positives are identified.

There are many different classification algorithms. Trees, rule-based models and probabilistic models have been employed successfully in many different machine learning domains, but have fallen short in BCI contexts (Lotte et al., 2018). Linear methods such as support vector machine (SVM), the aforementioned linear discriminant analysis (LDA) and other kernel methods have been the prominent tools employed in BCI classification. Perhaps the greatest reason for this is that SVM and LDA emphasize dimensionality reduction, and are optimized for such tasks. We have already mentioned how LDA reduces dimensionality by identifying the most relevant features. SVM similarly reduces dimensionality, but does so by constructing a hyperplane in high dimensional space. It then employs a kernel to classify the data. Though there are additional variants on these methods, the core concepts of dimensionality reduction, classification and test accuracy form the basis of machine learning for BCIs, and most of the attention in this subject has been paid specifically to SVM and LDA classification techniques (Quitadamo et al., 2017).

There are many tools that can be employed to perform this type of machine learning analysis, but not many of them are optimized for processing EEG data. Recently, a stack of Python-based tools have been developed which is optimized for this process. We have already mentioned that MNE Python facilitates the processing of electroencephalography (EEG) data in a Python based environment (Gramfort et al., 2013, 2014). One advantage of the MNE data structure is that it is compatible with other Python libraries that are commonly used in statistical and machine learning analysis such as NumPy (Van Der Walt, Colbert, & Varoquaux, 2011), Pandas (McKinney, 2010) and Scikit-Learn (Pedregosa et al., 2011). When combined, these tools form a single software environment that facilitates traditional ERP research, machine learning analysis, and BCI development. We will explore this further in Chapter 3.

2.4.3 Decoding Attention-Related EEG Signals

There has been considerable interest in decoding real-time signals over the past 20 years; this dissertation is by no means the first attempt to use these techniques. Some of the literature on this subject comes from aforementioned brain-computer intferface (BCI)s. Broadly, BCIs can be categorized into two categories based on data domains: frequency domain BCIs, which are most common and leverage band power or related parameters, and time domain BCIs which are less common and leverage time point parameters (Lotte et al., 2018). In the former category, motor-image BCIs have been widely studied. These have largely used LDA and similar techniques on band frequencies to perform classification. In addition, mental workload BCIs have been investigated, particularly by Information Systems researchers, which similarly use band frequencies to measure cognitive load (Pope, Bogart, & Bartolome, 1995; Berka et al., 2004), and have even recently been applied to the MOOC context (Lin & Kao, 2018). These BCIs are not typically used to measure specific attention processes in real-time, but can be used to observe patterns that are correlated with attention states.

The most prominent example of an attention-related BCI is the P300 BCI (henceforth expressed as P3 BCI), which was the first prominent BCI and similarly used an oddball paradigm to measure an attention-related phenomenon (Farwell & Donchin, 1988; Donchin, Spencer, & Wijesinghe, 2000). The P3 component is a well-studied event-related potential triggered through an oddball paradigm, and can also be evoked using an auditory stimulus. The P3 is often observed in conjunction with the P1-N1-P2 complex, but is only when participants attend to stimuli that are directly related to their task. It also exhibits a much larger amplitude than the complex, and is therefore much easier to detect. The earliest iteration of the P3 BCI was demonstrated by Farwell and Donchin in 1988, who designed a spelling interface. Using the P3 oddball effect, the characters of the alphabet are presented on a computer screen in a 6 x 6 matrix while the program flashes columns and rows randomly. The participant is asked to attend to a specific character, and when the character is evoked by the matrix, a P3 response is observed. After using a classifier, they created a process by which participants could spell words at the rate of 2.3 characters per minute.

Subsequent work on the P3 speller have focused on improving the tool's accuracy, or to applying it to other paradigms. The original speller as proposed by Farwell and Donchin (1988) used a stepwise linear discriminant classifier to identify P3 oddballs on crude epochs. Techniques such as SVM and LDA have been consistently demonstrated to perform well at these tasks, though a number of techniques have been explored recently which improve on them (Thulasidas, Guan, & Wu, 2006; Guger et al., 2009). Recent research has explored the application of random forests to improve the speed of P3 BCIs (Akram, Han, & Kim, 2015), the application of transfer learning to improve performance (Gayraud, Rakotomamonjy, & Clerc, 2017), as well as the use of convolutional neural networks (Cecotti & Graser, 2011). Though these techniques have been demonstrated to improve the P3 speller performance, they largely offer a marginal improvement of the traditional LDA/SVM classifier.

In addition to improving the original P3 speller, there have been efforts to apply the tool to novel applications. The P3 speller has been notably been applied to video games such as checkers; checkers movement selections can be iterated across a grid similarly to letters in the P3 speller (Fazel-Rezai et al., 2012). An open source iteration of the classic video game Space Invaders, called "Brain Invaders" has been developed using this paradigm, and released publicly (Congedo et al., 2011). Using the grid format similar to the speller, users focus on a particular point they would like to "destroy", using graphics similar to the classic game. A P3 brain-computer intferface (BCI) has been demonstrated in a virtual reality environment, which contained a number of objects (Edlinger, Holzner, & Guger, 2011). In this context, users were instructed to focus on particular objects to "interact" with them. Novel applications of BCI technology are an area of promising research in its own right.

A critical component of the P3 BCI is that, in addition to offline classification, they also employ some online signal detection method to enhance or transform the signal. The P3 component is sensitive to repetition effects, and online recordings are susceptible to blinks and muscle movements (Fazel-Rezai et al., 2012). In response, many online BCI classifiers employ an automated method to prepare the data, such as the xDAWN method explored earlier, which uses supervised learning to build an algorithm that maximizes the signal to signal+noise ratio (Rivet et al., 2009). By a method such as xDAWN, the classifier can interpret data that is similar enough to that which would be observed offline.

Finally, there is an emerging trend in P3 BCIs to emphasize hybrid effects, or incorporate the frequency domain. P3 classifier performance has been demonstrated to improve by incorporating frequency band activity (Speier, Fried, & Pouratian, 2013;

Käthner, Wriessnegger, Müller-Putz, Kübler, & Halder, 2014). This same activity can be used to incorporate other features to the BCI, such as mental workload effects. There are potential applications to leverage these indicators to identify patterns associated with attention-related constructs, such as mind wandering. In the future, it could likewise be used to build robust classifiers that explore different dimensions of the attention construct in real-time. This could potentially lead to a method for detecting variations of attention in real time for IS research and be further expanded to create dynamic interfaces or IT artifacts which adapt to users' attentional states.

2.5 Summary

In this chapter, we explored background literature from four subjects: learning theory, the neuroscience of attention, information systems, and machine learning for EEG analysis. We have discussed how attention plays a critical role in many of the major learning theories that inform MOOC design. However, measuring attention is not simple, as neuroscience informs us that it is actually an emergent process of at least three sub-processes: *alerting*, *orienting* and *executive control*. The information systems discipline has traditionally ignored the distinction between these sub processes, but has recently been motivated to explore the attention phenomenon using neurophysiology. Electroencephalography is a great candidate tool because it can not only measure these sub-processes, but also shows prospects for measuring attention states in real-time. In Chapter 3, we will describe the detailed hypotheses and theoretical approach that guides this dissertation, and ultimately how an improved, passive attention measure can be used to inform MOOC design. We will later draw from many of the concepts discussed in this chapter, but will either go into greater detail or will expand on some of the material mentioned here.

Chapter 3

Research Methodology

In Chapter 2 we explored the background of this dissertation's research question and ultimately concluded that a study of the neurophysiology of attention can help explain its role in effective e-learning. However, we are now faced with the challenge of bridging research methodologies across four disciplines. In this chapter, we will discuss how we will bridge methods in education, neuroscience, information systems and machine learning to conduct this research. We will start by developing the hypotheses of this research before proceeding to explore the appropriate methodologies for corroborating them empirically. We will then conclude that two experiments must be conducted in order to lend sufficient empirical evidence for the research contributions. These experiments and their results are subsequently discussed in Chapters 4 and Chapter 5.

3.1 Hypotheses and Research Model

When explaining hypotheses, information systems researchers often provide a model that can easily explain the different constructs and concepts explored. Though some concepts are excluded, we provide a model that will make many of the concepts easier to follow. Figure 3.1 is a model of the research questions described in this dissertation. We will explain each of these hypotheses in turn.

3.1.1 The Mind Wandering Experience

To understand the relationship between mind wandering and e-learning efficacy, we first need to develop a method for measuring mind wandering in the first place. Sullivan et al. (2015) developed an ex-post questionnaire and investigated the role of perceived mind wandering in an IS setting. As we have already seen, this is not sufficient for understanding the nuances of mind wandering. There is also a crucial difference between the ex-post questionnaire they developed and many of the other

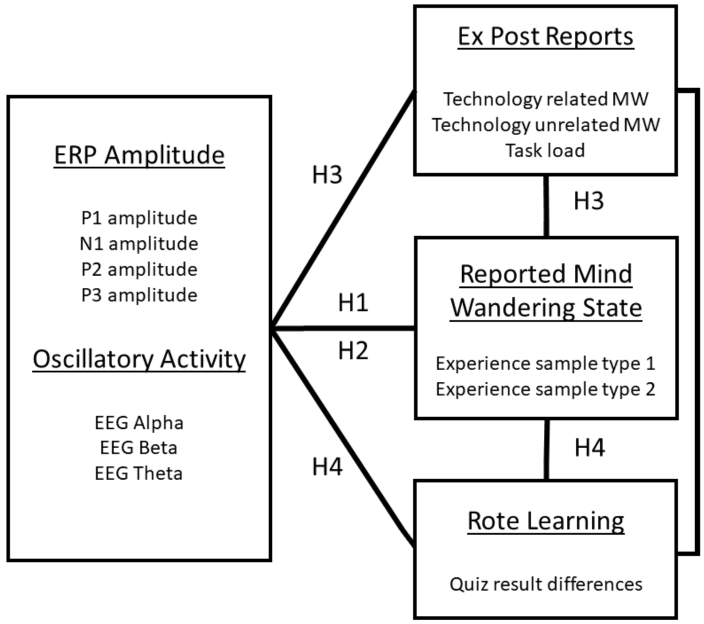


Figure 3.1: The proposed research model, excluding H5

self-report measures they drew from. They give examples of experimental studies that used questionnaire instruments to explore the relationship between subjective experience and observed attentional lapses (Smallwood et al., 2004; McVay, Kane, & Kwapil, 2009). However, these methods are different from those employed by Sullivan et al. (2015) insofar as they measure a present experience, not participants' perception of it afterwards. It is entirely possible that there are significant differences between mind wandering experienced during an attentional lapse and the perceived degree of mind wandering reported following a study.

Studies have explored the temporal element of mind wandering using a method called experience sampling. As previously mentioned, Lindquist and McLearn (2011) examined the impact of daydreaming and the lecture experience. They explored this relationship by using an audio probe, and asking participants to record whether they were experiencing task-unrelated thoughts at that particular moment (Lindquist & McLean, 2011). With this method, they were also able to determine that mind wandering was less likely to occur in the front third of a lecture hall, and also discovered a negative correlation between mind wandering and note taking. Similarly, Wammes and Smilek (2017) employed an auditory probe method to compare online and in-person lectures. They found different temporal patterns of mind wandering between electronic and in-person lectures, ultimately concluding that mind wandering increases when online, and does not generally increase when lectures are conducted in person.

One of the primary outcomes of this dissertation is to identify neurophysiological patterns that are correlated with mind wandering at a particular time, which we will henceforth refer to as the "mind wandering state". Such experience samples are good correlates of EEG patterns, because they give us a fixed time point to compare to. There is an implicit problem with doing this however; experience samples are disruptive. One of the primary motivations for building EEG measures in the first place is to develop a non-disruptive method for measuring mind wandering. By correlating EEG patterns to experience samples directly, we would run the risk of not genuinely measuring the mind wandering state.

One way to get around this is to observe the time segments immediately preceding an experience sample probe. Braboszcz and Delorme (2011) took this approach, and observed the 10 seconds before a mind wandering report with the 10 seconds after. In that study, participants performed a meditation task while auditory tones were played. They observed the P2 event-related potential (ERP) triggered by infrequent oddball paradigm auditory stimuli and found differences in the amplitudes between the mind wandering (before a report) and breath focus (after a report) conditions. Though their findings emphasized observed differences in the P2 specifically, they also observed differences in P1 and N1 amplitudes. Furthermore, they observed differences in oscillatory patterns. Low frequency patterns (less than 13 Hz) were found to be correlated with the mind wandering states while high frequency patterns were inversely correlated. Following Braboszcz and Delorme (2011), we can similarly hypothesize that we would see these differences in patterns in an e-learning setting, though we should consider excluding delta band analysis (1-4 Hz) due to the risk of measuring blinks, which are often detected at that frequency.

An alternative hypotheses can be made about the P3 component. A past study by Smallwood et al. (2008) identified a reduction in the P3 component that is associated with mind wandering during a sustained target detection task (Smallwood et al., 2008). Reductions in P3 amplitude were reflective of mind wandering and participants diverting their attention to internal task-unrelated thoughts. Similarly, we can expect a similar effect from a sustained attention task such as learning from a video. In an auditory paradigm such as that employed by Braboszcz and Delorme (2011) however, we could also posit that P3 amplitudes would increase during states of mind wandering, as participants' attention is directed away from the task of learning from the video and towards the auditory sounds. These alternative explanations are summarized in Figure 3.1 by H1 and H2 and their sub-hypotheses, where the "reported mind wandering state" is used to denote in-session self reports. The hypotheses are summarized below.

H1a — Mean P1 amplitude will distinguish on task and mind wandering states.

H1b — Mean N1 amplitude will distinguish on task and mind wandering states.

H1c — Mean P2 amplitude will distinguish on task and mind wandering states.

H1d — Mean P3 amplitude will distinguish on task and mind wandering states.

H1d — Mean theta power will be positively correlated with the reported mind wandering state.

H1e — Mean alpha power will be positively correlated with the reported mind wandering state.

H1f — Mean beta power will be negatively correlated with the reported mind wandering state.

An important limitation to these hypotheses is the difference in experience sampling method used by Braboszcz and Delorme (2011) and the other literature discussed in this section. During a meditation task, participants were asked to conduct a specific breath counting exercise, and were asked to report when they began to lose count of their breathing. In an online lecture context, there is no clear task by which this distinction can be made, and we would rely on participants' subjective experience of what constitutes sufficient mind wandering during an online lecture task. An alternative approach involves a pseudo-random interval interruption, when participants are asked about whether they experienced mind wandering. Using a Likert scale at these points, we can mitigate this problem of subjective threshold (Wammes

& Smilek, 2017). It is unclear which of the two methods is more appropriate in the online lecture setting, so we propose exploring both. In order to do this, we will need to conduct two studies, one with the experience sampling method described by Braboszcz and Delorme (2011) and another with the method described by Wammes and Smileck (2017). We are thus led to hypothesize that there will be no significant difference between the two.

H2 — There will be statistically significant correlations between neurophysiological measures and both types of experience samples.

Sullivan et al. (2015) found evidence to support that mind wandering or on-task thought could be characterized in terms of technology (task) related and non-technology (task) related when discussed in an information technology setting. They found that on-task thought which was technology-related was positively associated with knowledge retention and that technology-related mind wandering had a negative impact (Sullivan et al., 2015). However, non-technology related mind wandering had no impact on knowledge retention. This work thus established a method for investigating mind wandering in an education setting which suggests a distinction that may be useful when similarly investigating the efficacy of MOOCs.

This said, the authors acknowledge significant limitations to their study. They recognize the problem of a missing time dimension, the lack of neurophysiological correlates and even call for a future investigation using neurophysiological tools. Additionally, they acknowledge that the study is limited insofar as it only tested the moderating effect of mind wandering, and ignored other potential moderators such as working memory capacity and task complexity (Sullivan et al., 2015). A robust extension of their work should not only consider neurophysiological correlates, but also the impact of other moderating factors.

A simple way to incorporate this is to include additional ex post questionnaire content related to working memory and task complexity. The NASA Task Load Index (TLX) incorporates working memory activation and task complexity, albeit subjectively (Hart & Staveland, 1988). The original TLX consists of six dimensions of task load: mental demand, physical demand, temporal demand, performance, effort and frustration, each answered using a subjective subscale ranging from low to high.

By incorporating the a reported task load measures in addition to the mind wandering reports, we can explore the relationship between the two constructs.

Assuming that certain neurophysiological indicators are associated with the mind wandering state, and that mind wandering can be reliably measured using an ex post measure, it makes sense that there would be a similar correlation between the ex post measure and the neurophysiological indicators. However, while the previous statements concerned particular states in time, the ex post measure concerns the period following the MOOC session; in other words, it tracks users' perception of their experience rather than the experience itself. The neurophysiological indicators in question must therefore represent an entire online lecture session, rather than small segments of time around experience samples. Assuming the mind wandering state and participants' perceptions of it are consistent, we are led draw hypotheses about the relationship between the ex post measures and session measures. Rather than specifying the relationship between each component separately, we will combine hypotheses about the individual components to hypotheses about component amplitudes generally.

H3a — Component amplitudes (P1, N1, P2, P3) will be positively correlated with reported Technology unrelated mind wandering.

H3b — Component amplitudes (P1, N1, P2, P3) will be positively correlated with reported Technology related mind wandering.

H3c — Component amplitudes (P1, N1, P2, P3) will be negatively correlated with reported task load.

H3d — Reported mind wandering experiences will be positively correlated with reported technology unrelated mind wandering.

H3e — Reported mind wandering experiences will be positively correlated with reported technology related mind wandering.

H3f — Reported mind wandering experiences will be negatively correlated with reported task load.

3.1.2 The Impact of Perceived Mind Wandering

Where hypotheses 1 through 3 concerned the first research question about measuring the mind wandering state, hypotheses 4 and 5 concern mind wandering's impact on learning. Assuming that we can develop a reliable measure of mind wandering, we can then apply the measure to develop insight into its impact on learning. The fundamental task of schooling is to facilitate learning, and online lectures or MOOCs are no exception. This is to say that the purpose of schooling is to facilitate positive learning outcomes. In Chapter 2, we explored the concept of learning and outlined a number of theories about how learning is facilitated either online or in a classroom environment. These theories ultimately explore methods for teaching and learning better. Yet, what are learning outcomes? A teacher may go their whole life without asking this fundamental question.

Mayer contrasts three kinds learning outcomes, and though he makes this assertion in the discussion about the Cognitive Theory of Multimedia Learning (CTML), these are applicable to classroom teaching in general (Mayer, 2009). The first outcome is *no learning*, where a student is unable to retain the content of a lesson, performs poorly on tests of retention, and is unable to demonstrate knowledge transfer. The second outcome is referred to as *rote learning*, where students are able to retain facts (eg. can remember steps in a process) but are unable to use this knowledge to solve problems creatively. Rote learning is often distinguished by good retention but poor knowledge transfer. The third outcome is what Mayer calls *meaningful learning*, also often referred to as inferential learning, where students not only retain knowledge, but demonstrate knowledge transfer, perhaps by using the knowledge to solve novel problems. All this is to say that knowledge retention precedes mastery.

This distinction is useful because it helps us establish tangible goals in an investigation. Except for extreme circumstances, we can expect a single electroencephalography (EEG) session to last up to two hours, after which participants would need to rest. Meaningful learning is therefore unlikely to be acquired within the limits of a single session. This said, it is entirely possible and reasonable to demonstrate rote learning within such a timeframe and it would be useful to constrain the type of learning conducted by a study. By constraining the study to a single session and the acquisition of rote learning outcomes, we can establish more detailed temporal analysis which can better inform the design of individual MOOC sessions. Given our previous discussion about how neurophysiological indicators are predicted to relate to mind wandering.

Tests are perhaps the most common measure of successful of learning outcomes and one of the most common methods of testing MOOC efficacy. An investigation into MOOC efficacy could incorporate tests, which many MOOCs already incorporate into their design. Multiple-choice testing, in particular, is pervasive in MOOCs and is often used as the primary means of student assessment in a university environment. The main motivation for multiple choice testing appears to be its scalability and ability to accommodate large class sizes (Roediger & Marsh, 2005) and not necessarily its efficacy. There has been growing concern about the applicability of multiple choice as a primary test mechanism, largely because of challenges with it incorporating different learning outcomes. For the measurement of rote learning however it is appropriate, and a simple multiple choice instrument could be successfully used for such a study. By having participants complete a test before participating in a lecture video, we can reliably measure their current understanding of the topic. By subtracting scores on the same test afterwards, a measure we will call *quiz deltas*, we can see what they learned from the video. Finally, we should also consider the relationship between the selected ex post measures and rote learning outcomes. Assuming that the mind wandering experience generalizes to ex post perceptions, we should also observe positive relationships between ex post reports and rote learning outcomes. Assuming that mind wandering has a negative impact on learning, we are led to the following set of hypotheses:

H4a — Quiz deltas will be negatively correlated with component amplitudes.

H4b — Quiz deltas will be negatively correlated with reported technology unrelated mind wandering.

H4c — Quiz deltas will be negatively correlated with reported technology related mind wandering.

H4d — Quiz deltas will be negatively correlated with reported task load.

H4e — Quiz deltas will be negatively correlated with reported mind wandering states.

3.1.3 Mind Wandering State Decoding

This brings us to the end of Figure 3.1. One final objective of this dissertation is to create a real-time measure of mind wandering that can be used in MOOC research, as well as other IS contexts. The experiment and data as described in this chapter will make it possible to create such a real-time measure. The question of how such a measure can be deemed successful should be addressed.

As mentioned earlier, though there do not appear to be any extant examples of decoding in this context, there is considerable research on other ERP detection tasks, such as that performed by P3 BCIs (Quitadamo et al., 2017). In the case of the P3 speller, performance of 96.5% accuracy (Rakotomamonjy & Guigue, 2008) using Ensemble SVM, 85% accuracy using least squares SVM (Gu, Yu, Shen, & Li, 2013) and 90% accuracy using standard RBF kernel SVM (Salvaris & Sepulveda, 2009) have been reported. These seem to be useful benchmarks, insofar as they involve a similar task and all use variants of the SVM classification algorithm. However, we should be cautious when using these as a benchmark for our task. The P3 is a substantially more powerful signal, with amplitudes often exceeding twice that of the P2 (Luck, 2014). A P2 time-series classifier is thus likely to perform considerably more poorly than these benchmarks given that it will have a much lower signal-to-noise ratio.

A better benchmark was given by Furdea et al. (2012), which explored an ERP BCI for a different task. In this work, they describe a brain-computer intferface (BCI) that classifies conditioned reactions to semantic stimuli, with the goal of eventually developing a BCI to give paralyzed individuals the ability to communicate verbally (Furdea et al., 2012). They constructed a paradigm for controlling either agreement or disagreement (”yes/no”) to short propositions, and classified the subsequent low-amplitude waveform. They compared multiple classification techniques and ultimately found that radial-basis function kernel SVM exhibited the best classification result at 68.6% accuracy. This was was not statistically significant and was not sufficient to conclude that classification was being performed above the chance level of the two reactions, given their sample size.

We can likewise expect low classification results, given that we will have a low amplitude waveform, relatively few samples, and considerable noise during recording. However, we also have reason to believe that it can be accomplished and that the results will be better than random chance. The P2 ERP has been studied extensively, and has been used to demonstrate clear differences between mind wandering and on task states. Further building on the work described in the BCI literature, we can also hypothesize that SVM or one of its derivatives will perform best at classification tasks (Lotte et al., 2018). This leads us to articulate our final hypothesis:

H5 — An algorithm can be constructed which measures differences in reported mind wandering states with at least 70% accuracy.

3.2 Development of Questionnaire Instruments

3.2.1 Pre and Post Study Questionnaires

Following the development of the conceptual framework, we investigated the IS literature for other suitable measurements of mind wandering. To the best of our knowledge, there is only one instrument that has been developed to investigate the role of mind wandering in the IS setting (Sullivan et al., 2015). Given that this is a post-hoc questionnaire measure, this is a natural candidate for our investigation into the relationship between traditional and neurophysiological measures, which is among the main reasons it was selected for this study.

Sullivan et al. (2015) developed their instrument to incorporate the relationship between on-task thought, mind wandering, creativity and knowledge retention. In addition to the four constructs measured, they further distinguished technology-related mind wandering and non-technology related mind wandering. Some of the instruments they developed are relevant to our inquiry. They explored the role of creativity on mind wandering because it has been separately found to be associated with the unconscious mind and the way that it processes ideas (Dijksterhuis & Meurs, 2006). Given that mind wandering is associated with unconscious processing, an investigation into the relationship between creativity and mind wandering was justified in their case, but not in ours. They also explored this technology/non-technology distinction because they hypothesized that they would result in different behavioural outcomes, especially with respect to the moderating effect of creativity. We are not interested in exploring the role of creativity, so we can exclude it.

In order to investigate their hypotheses, Sullivan et al. (2015) developed questionnaire measures for each of the four constructs, in addition to a number of control variables. Most of the instruments were derived from previously validated measures, though some were newly introduced. To validate their scale, they conducted a pretest of the items, and then conducted a full test of their structural model. The result was in a valid mind wandering instrument which they could use to test their hypotheses. They eventually found that mind wandering did have a moderating effect on both creativity and knowledge retention, but also that this relationship depended on whether the mind wandering was related to the information technology task. In our case, we can leverage the technology-related and non-technology related mind wandering scale to have a valid post-hoc measure.

As previously mentioned however, Sullivan et al. (2015) ultimately concluded that there are some significant limitations to their study, one of which was its neglect for the role of working memory capacity and cognitive load (Sullivan et al., 2015). This is why we complement their measure with questions derived from the NASA Task Load Index (TLX). Though there are other measures of cognitive load, some of which are specifically tailored to Sweller's model or multimedia learning (Brunken, Plass, & Leutner, 2003; Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013), none are as widely used and accepted as the NASA TLX and few of them incorporate extraneous load from a task, such that caused by sitting still through a long electroencephalography (EEG) experiment. Furthermore, the questionnaire format of the TLX makes it practical to implement in a ex post fashion. The NASA TLX therefore represents a very well validated measure with which we can confidently benchmark physiological measures against.

As such, we constructed a 25 question post-study measure and pre/post test measure of rote learning. The 25 questions were derived from Sullivan et al. and the NASA TLX, and were designed to be answered on a 7-point Likert scale. The pre/post test consists of 10 multiple choice questions on content from the video. By creating a pre/post measure, we can effectively account for knowledge that participants had before joining the experiment while also accounting for rote learning and the difficulty of subject matter. The detailed pre/post test measures are provided in Appendix B while the post-study questionnaire is detailed in Appendix C.

3.2.2 Experience Sampling Methods

In addition to pre/post measures, we can also conduct experience sampling. In the hypothesis development subsection we explore two sampling methods: one where participants subjectively assert when they experience mind wandering (Braboszcz & Delorme, 2011) and one where they are prompted to report on their mind wandering experience (Wammes & Smilek, 2017). In the first has the advantage of being able to neatly distinguish when participants experience wandering from when they are not. It also has the disadvantage of not accounting for the degree of experienced mind wandering, or for the variance in subjective experiences throughout the lecture. The later has the advantage of being able to account for this variance.

In order to account for the potential differences in these methods, we propose conducting studies that explore each of them. In the first study, participants can report when they experience escaping the mind wandering state. These samples can be aggregated and used as either an ordinal or continuous variable in regression analysis, and used to distinguish conditions for EEG analysis. In the second study, the fixed prompt method can be employed. These samples could be again used in regression to compare continuous variables. The subjective measures can also be discretized for comparison.

Appendix D describes the two sampling methods. Method 1 describes the self-reported mind wandering experience. Method 2 describes the potential subjective responses to the mind wandering prompt. The methods can be further compared by whether we see statistically significant differences in explanatory power between the experience sampling methods used in the two studies.

3.3 How to Conduct and Analyze an Auditory Oddball Experiment

There has been some discussion in the NeuroIS community about the appropriate technical methods for conducting neurophysiological research. While some researchers

Tool Name	Description
Anaconda	A distribution of the Python and R programming language for scientific computing (Anaconda, 2019).
Jupyter	A notebook format for sharing code and computational narratives (Perez & Granger, 2015).
Matplotlib	A 2D graphics package for the creation of publication-quality images (Hunter, 2007).
Numpy	A package for scientific computing and analysis (Van Der Walt et al., 2011).
Pandas	A data library optimized for manipulating large and time series data (McKinney, 2010).
PsychoPy	An application and library used to run psychology and neuroscience experiments (Peirce, 2007).
MNE	A library for preparing, analyzing and visualizing MEG, EEG and other related data (Gramfort et al., 2013, 2014).
Scikit Learn	A machine learning library for Python (Pedregosa et al., 2011).

Table 3.1: Summary of Python tools used for EEG analysis

have been motivated to develop proprietary platforms (Courtemanche et al., 2018) or open source libraries (Michalczyk, Jung, Nadj, Knierim, & Rissler, 2019), there is an alternative approach which can be taken. Instead of developing new tools, existing tools can be adapted to conduct complex experiments. In this subsection, the tools and approach used to develop experiments and analyze data are described. All of the software tools discussed are written in the Python programming language and are freely available under the Python Software Foundation's open source license. The "Python Stack" subsequently described can likewise be adapted for other EEG experiments and it is our hope that this section will be useful to researchers who wish to explore similar questions in the future.

3.3.1 Python Tools for EEG Processing

In the experiments described in this dissertation, the Python programming language was used to deliver stimuli, collect data, analyze data and conduct machine learning analysis. Table 3.1 summarizes the tools used throughout the experiment. Though the Python programming language, like all programming languages, requires technical capabilities to use, it is highly portable and is designed to be readable by a broad audience. Specialized data science Python packages such as Anaconda (Anaconda,

2019) come with robust package managers, as does Python itself. Package managers make it relatively simple to install and configure the packages necessary to conduct and analyze the experiment.

We encourage the use of notebooks when conducting analysis for scientific experiments. Notebooks such as Jupyter (Perez & Granger, 2015) facilitate collaboration by combining code and markup. They also help make scientific analysis reproducible by forcing scientists to record their scripts and transparently present results (Toelch & Ostwald, 2018; Harding, 2019). This is particularly useful when conducting EEG analysis because of the amount of subjective preprocessing which is often conducted on participants. Ineffective preprocessing can lead to many problems, and has been raised as a potential methodological concern in the event-related potential community (Luck & Gaspelin, 2017). Replication of preprocessing analysis and data transparency can help neuroscientists and NeuroIS researchers manage this problem.

3.3.2 Design of Auditory Oddball Stimuli

In Chapter 2 we discussed the auditory oddball experimental paradigm. In that experiment, auditory stimuli are presented which elicits the P1-N1-P2 and P3 response. In the standard oddball paradigm, 80 percent of the stimuli are presented at one auditory frequency while 20 percent are presented at a noticeably different frequency. The oddball stimuli (20%) elicit a response that is more pronounced, but also sensitive to differences in covert attention.

To the best of our knowledge, we are the first to apply this paradigm to an adult MOOC context. We conducted experiments where the standard auditory oddball paradigm was enacted, but where participants are also asked to observe MOOCs. As shown by Braboszcz and Delorme, the oddball stimulus response is different in the 10 seconds before a mind wandering report from the 10 seconds afterwards (Braboszcz & Delorme, 2011). One of the ways that we observed mind wandering is by comparing the neurophysiological responses throughout the online lecture experiment. In an effort to replicate their findings in a different setting, we can likewise observe the 10 seconds before and after a button press. Figure 3.2 demonstrates the generic setup for our experiments.

EEG recording was conducted over the length of a video as oddball stimuli are

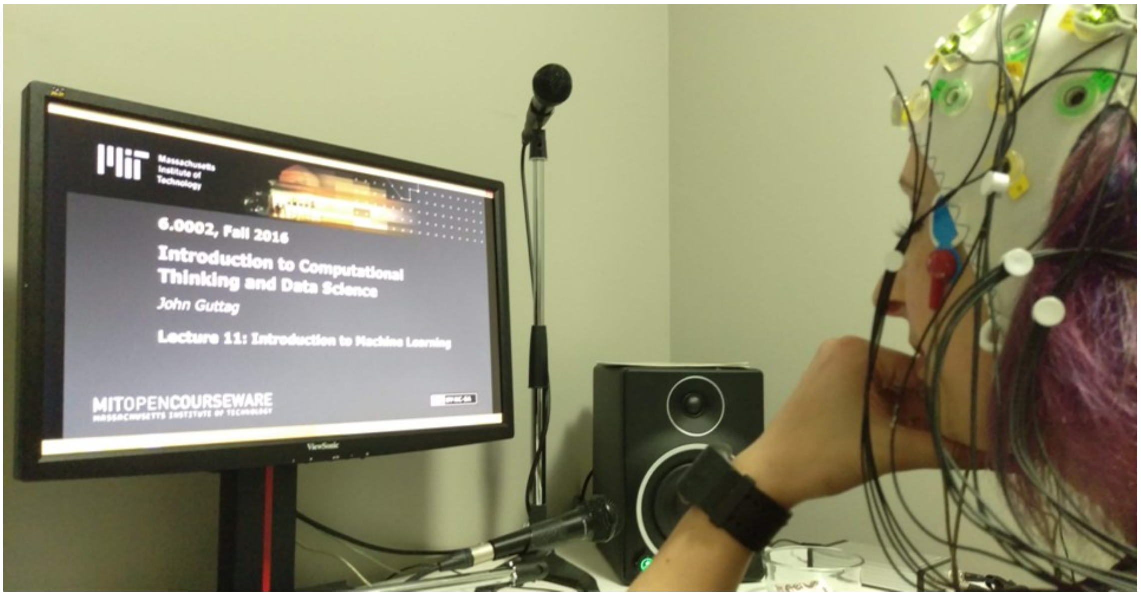


Figure 3.2: A demonstration of the proposed experimental setup

presented. Oddball stimuli were developed using the PsychoPy platform which comes with auditory stimuli objects (Peirce, 2007). The timing of the auditory signals was recorded by PsychoPy using its pre-defined parallel function, which allowed us to activate pins on the computer's parallel port. The result is a signal that can be read in the live EEG data. PsychoPy was also used to administer surveys and mind wandering prompts. PsychoPy also helped us streamline the aggregation of the data generated throughout the section (i.e. timings and survey results) by simplifying each recording in a single csv file.

According to Luck (2014) we require approximately 400 recordings to accurately compare the P1 component; we therefore require long video sessions during which participants' electroencephalography (EEG) data was recorded. We employed a standard 32 channel EEG system with a modified montage to emphasize the paretal and occipital regions of the head, given that this is the region is where the P1-N1-P2 signal is strongest (Hillyard & Anllo-Vento, 1998). By using the actiCHamp EEG system (BrainVision LLC) with low impedances, we can collect high quality signals even with a limited number of channels.

3.3.3 Preprocessing, Epochs, Artifacts and Analysis

Once the data was collected, it needed to be prepared. Using MNE (Gramfort et al., 2013, 2014) users are able to extract prepare data from supported devices. EEG data can then be saved as a *mne.raw* object which can be subjected to processing. Raw objects can be subjected to bandpass filtering, which eliminates noise either below of above a defined threshold. The *mne.preprocessing* library provides a number of methods that can be further applied to this step such as the xDAWN algorithm, which eliminates artifacts and amplifies differences between ERP conditions (Rivet et al., 2009).

ERP experiments require researchers to analyze data which is time-locked to events rather than a continuous stream of data. MNE provides an *mne.epochs* data format which is well-suited to this task. This data type can be generated from the *mne.raw* data based on the events recorded in the eeg data itself. The *mne.epochs* objects which are generated can then be visualized and inspected using a number of helpful visualization tools that are written into the objects as methods. The visualizations are built on Python's matplotlib library (Hunter, 2007) which is designed to create publication quality visualizations.

In the experiments described in this dissertation, we expected there to be considerable noise created by blinks and head movement throughout the session, which have to be filtered before analyzing the data. We used the visualization method described above to manually investigate the epochs. We then applied MNE's independent components analysis (ICA) library to further clean the data. ICA separates the epoch data into independent components which explain the variance of the data. When data are recorded correctly, ICA components which are the result of EEG artifacts will be visually apparent. Using this method, we can therefore select and remove components that are likely artifacts (Delorme & Makeig, 2004). We can then save the code that cleans the futures so that the precise processing results can be replicated.

Following preparation, the oddball response can be averaged compared visually. We can identify the timing of the the amplitudes and peaks of the P1, N1, P2 and P3 components and compare the responses from each of the conditions during these time windows. We compared the amplitudes of four conditions: two standard conditions (mind wander and on-task) and two oddball conditions at the three identified time

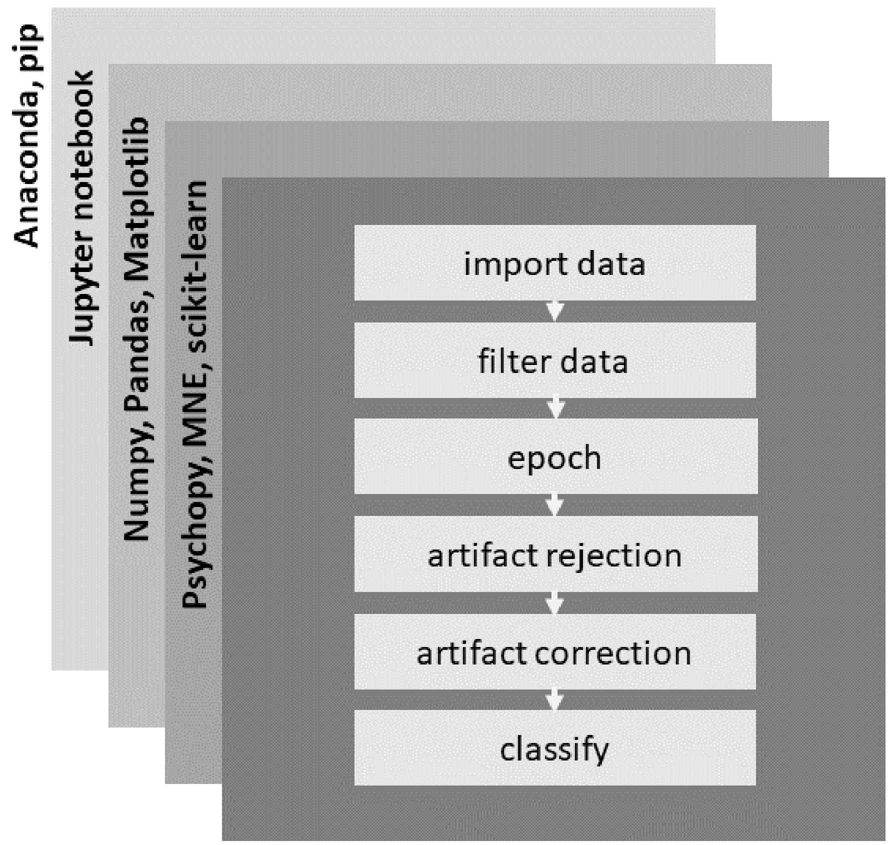


Figure 3.3: An illustration of the Python Stack and EEG data analysis process

windows. We also identified relevant electrode regions to compare by first visualizing the average differences during each time window and then extracting a region of interest to be compared consistently across participants.

In addition to the event-related potentials, we can also compare oscillatory activity. As mentioned in Chapter 2, unlike ERP, oscillatory activity is often contrasted in the frequency domain at various bins commonly discussed in the literature: delta (0-3 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma (30-100 Hz). When contrasting in the frequency domain, power spectrum analysis is commonly employed. Frequency power, the squared oscillation amplitude at a given frequency band, can be calculated across a given period of time. Power spectrum analysis can also be compared statistically across specific electroencephalography (EEG) channels, much the same way as analysis on ERP amplitudes. In our analysis, we can calculate the power at the theta, alpha and beta bands during the 10 seconds before a mind wandering or on-task report and compare the results.

To quickly summarize, ERP comparisons were made between four conditions: reported mind wandering standard responses, reported mind wandering oddball responses, reported on-task standard responses, and reported on-task oddball responses. Following Braboszcz and Delorme (2011), statistical comparisons will be made among the responses from the 10 seconds before a report. ERP components will be identified visually by observing averages, segmented by selecting specific time windows and groups of electrodes, and then compared. Likewise, oscillatory activity will be observed at a collection of electrodes in the 10 second windows before a reported state. We shall now discuss the statistical methods that will be employed and how they will help maintain validity in the results.

3.3.4 Techniques for Ensuring the Validity of EEG Results

The methodological challenges of statistical analysis for EEG were articulated in an article eloquently titled "How to get statistically significant effects in any ERP experiment" (Luck & Gaspelin, 2017). In this article, they highlight a fundamental challenge with ERP experiments. Given the massive size and high dimensionality in the data, it is likely that *any* given ERP experiment will produce statistically significant results of some sort. A fundamental issue is that ERP component amplitudes windows are averaged, visualized, and chosen arbitrarily. In doing this, experimenters are implicitly making hundreds of statistical comparisons (Luck & Gaspelin, 2017). Further, given that multifactor statistical analysis is frequently used in ERP studies, it is likely that we will observe false positives. The experiment described thus far is no exception to these criticisms, as we use the same data preparation process as that critiqued by Luck and Gaspelin (2017).

This dissertation incorporates two methods to manage these concerns. The first is that we use mixed effects analysis to interpret the results (Baayen, Davidson, & Bates, 2008). Mixed effects analysis can be used to include both fixed effects (nonrandom and independent effects) and random effects, which is helpful for contexts with repeated measures within subjects. It allows us to incorporate both within-subject effects and intra-electrode effects in our analysis, in addition to the between-condition effects we desire observe. LME helps mitigate many of the risks of accidentally detecting between-subject effects which Luck and Gaspelin (2017) warn about by accounting for random effects which may be present in the data. It has not yet been widely adopted by NeuroIS researchers, though linear mixed effects analysis was recently employed in a study of habituation in security warnings (Vance, Jenkins, Anderson, Bjornn, & Kirwan, 2018). One challenge with this method is that it assumes a linear relationship, which may not be the case. We will employ a generalized additive mixed effects model (Tremblay & Newman, 2015) we might ensure that our results accurately account for variances between participants, especially given the high degree of expected variance in responses.

The second, and perhaps most important consideration is experiment replication. Luck and Gaspelin (2017) recommend replication in any experiment where independent variables are arbitrarily selected. In our experiment, we must select component time windows and the period of time to select epochs to compare. Our study could therefore be a strong candidate for type 1 errors. We manage this by attempting to replicate the results described in Braboszcz and Delorme (2011) in the online lecture context. We further manage this by conducting two experiments and cross-referencing the results with two different mind wandering report techniques, which helps ensure that significant results discovered are generalizable.

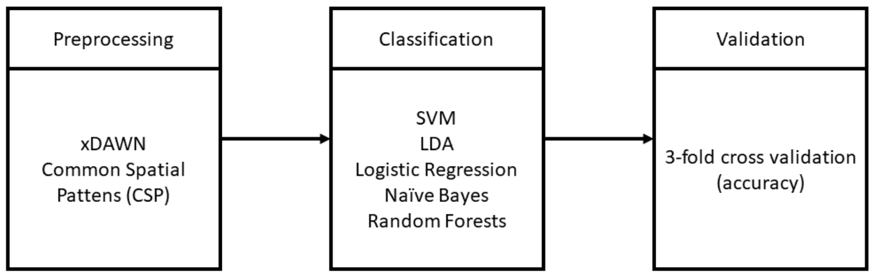


Figure 3.4: An illustration of the decoding process used

3.3.5 Techniques for EEG Decoding

As previously mentioned, there are two broad categories of decoding techniques and brain-computer intferface (BCI)s which are relevant to this dissertation. The classifiers employed in this dissertation each involve three steps: preprocessing, classification and validation. We shall describe the techniques used in each of the three steps for both the frequency domain and time-domains. This classifier observes frequency data from the 10 seconds before a report and applies three steps. Likewise, the time-series classifier assesses 1.2 second epochs around the auditory signals and applies preprocessing, classification and validation. Both of the tasks will be further explored in aggregate and for each participant individually.

Concerning preprocessing, the technique employed will depend on the domain being observed. Common spatial pattern filters are often used in frequency domain tasks and will be employed in this analysis (Koles, 1991). Common spatial patterns identify the frequency domain differences between two classes decomposes the data into subcomponents that maximize the differences. Time series data, by contrast, will apply the previously discussed xDAWN algorithm, which maximizes the signal to signal+noise ratio between the epochs. It is important to note that MNE Python contains libraries that simplifies these steps, which will be used to perform this analysis (Gramfort et al., 2013, 2014).

Following preprocessing, data can be classified using any number machine learning techniques. As previously discussed, given the dimensionality of the data, we expect two of these techniques to perform well: support vector machines and linear discriminant analysis. In addition, we shall explore three other classification techniques: Logistic Regression, Naive Bayes, and Random Forests for control. Each of these techniques are implemented in the Python scikit-learn library (Pedregosa et al., 2011). These techniques have been explored in previous BCI literature and are often employed in standard machine learning classification analysis (Lotte et al., 2018; Quitadamo et al., 2017).

Finally, validation shall be performed using a shuffled 3-fold cross validation (Kohavi, 1995). K-fold cross validation is a non-exhaustive cross validation technique where the data is divided into three folds. The classifier is trained on two of the folds and tested on the third; the process then continues for the remaining folds. Accuracy, as well as other desired measures, are measured for each fold-task. Scores are then averaged. This method is often employed in machine learning tasks with datasets with few samples, such as the classification task before us.

3.4 Summary

The hypotheses and methodology used in this dissertation were described. We discussed the theoretical model and the six categories of hypotheses to be corroborated. We also discussed the questionnaire methods employed, and how proven mind wandering and task load survey measures will be used to validate the neurophysiological measures. Components will be observed visually and compared at consistent time windows at each of the components, while power will be compared at the theta, alpha and beta frequency bands. We discussed how mixed effects modelling can account for differences among participants and electrodes, while incorporating experimental replication will help ensure the generalizability of our results. Decoding can also be performed using MNE Python, which may yield a real-time mind wandering measure. In chapters 4 and 5 we discuss two experiments that both draw from the methodologies described in this chapter. We shall begin with a smaller study which attempts to replicate the results Braboszcz and Delorme (2011) in an online lecture setting.

Chapter 4

An Experiment to Measure Mind Wandering during an Online Lecture

As mentioned in Chapter 3, prior literature has found that differences in early responses to auditory evoked potentials have been found to be useful indicators of mind wandering. Braboszcz and Delorme (2011) observed that differences in the P2 auditory evoked potential were significantly different between states of reported mind wandering and on-task experiences. In that study, participants were asked to perform a meditation task and press a button when experiencing mind wandering. Auditory stimuli were presented following an oddball paradigm and responses to the auditory tones were found to be significant predictors of mind wandering. However, the task described by Braboszcz and Delorme (2011) has some challenges that prevent their findings from being generalizable in an online lecture setting. Not least, they asked participants to perform counting task which gave a precise method of determining when mind wandering occurs. Such a counting task would inhibit performance in a secondary task such as learning from a video.

In this chapter we describe the results of a study that sought to replicate some of the key findings of Braboszcz and Delorme (2011). We are motivated to replicate their study to determine whether their findings are generalizable in an online lecture or broader e-learning context. By doing this, we make strides to identify the neurophysiology of the mind wandering state. This forms the foundation for a larger study which can validate the discovered measures with other methods, including those with lower expected effect sizes such as questionnaires, subsequently described in Chapter 5.

The primary goal of this study was to identify neurophysiological patterns that are correlated with mind wandering at a particular time. We predict that the P1, N1 and P2 responses to oddball auditory stimuli will be larger when participants report being in a mind wandering state than when they are on-task. We believe that this

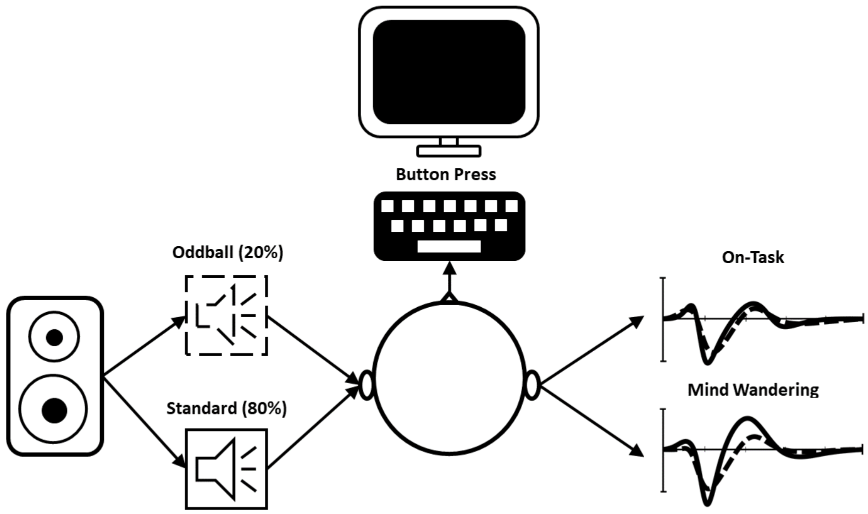


Figure 4.1: An illustration of the experiment design described

will be the case because when attention is directed at sensory stimuli, such as the auditory sounds, the amplitudes of the P1, N1, and P2 cortical responses to those auditory tones will be more pronounced.

The second objective of this study is to determine whether we can accurately detect mind wandering instances during an online lecture. We employed the previously described decoding processes using both time domain and frequency domain data. By doing so we identified an appropriate decoding method which can be applied to a larger population sample using a different experience sampling method.

To identify the moments when mind wandering occurred, we use the experience sampling method reported in Braboszcz and Delorme (2011). Participants were asked to press a button when they realized that their mind was wandering. Participants were told that they would be tested on the content and were tested following We likewise analyzed responses from the 10 seconds preceding the mind wandering responses and the 10 seconds afterwards.

4.1 Methods

4.1.1 Participants

Sixteen healthy students (7 females and 9 males; age 19-29 years, mean 23.6 and SD = 3.7) gave written consent to participate in the experiment. Participants were excluded if they reported neurophysiological, emotional, medical, hearing and vision conditions that could lead to abnormal electroencephalography (EEG). Participants were also excluded if they were majoring in computer science or had taken a course related to machine learning (which could confound the results of the study, which used an instructional video on this topic), or were not fluent in English (the language of the instructional video). Participants provided written and informed consent and were financially compensated CAD $25 for their time. All procedures were reviewed by Dalhousie University's Research Ethics Board. As detailed below, 6 of the 16 participants were excluded for either technical reasons or because they did not report enough mind wandering episodes.

4.1.2 Experiment Stimuli

Audio stimuli consisted of 100 ms tones; the frequent standard tones were 500 Hz while the infrequent oddball tones were 1000 Hz. In total 2446 tones were presented throughout the experiment. Oddballs and standard conditions were assigned to each tone randomly at a 80:20 ratio.

4.1.3 Procedure

After completing the informed consent procedure, participants were fitted with the electroencephalography (EEG) cap (see next section) and brought to the testing room. Participants were asked to attend to a 51-minute English language video on the subject of machine learning (Grimson, 2017). The auditory tones were presented at the same time as the video, over the lecture audio. The subject matter and video were chosen because it had some utility to the participants. Pilot testing suggested that this video would trigger variations in mind wandering and alertness for most participants. The video consists of a lecturer talking, along with an occasional visual aid created in Microsoft Power Point. Participants were asked to pay attention to the video and ignore the audio stimuli, which were presented at random every 1-1.5 seconds (mean 1.25). As in Braboszcz and Delorme (2011), participants were asked to report when they experienced mind wandering by pushing a button on the computer keyboard. Participants were also asked to complete a multiple choice quiz which was administered before the video and again following the video, in an effort to control for task difficulty. The PsychoPy library was used to present the audio stimuli and record manual responses (Peirce, 2007).

4.1.4 EEG Recording

Participants were fitted with 32-channel scalp electrodes (ActiCap, BrainProducts GmbH, Munich, Germany) positioned at standard locations according to the International 10-10 system and referenced during recording to the midline frontal (FCz) location. Bipolar recordings are made between the outer canthi of the two eyes, and above and below one eye, to monitor for eye movements and blinks. Electrode impedances were kept below 15 kOhm throughout the experiment. EEG data are sampled at 512 Hz using a Refa8 amplifier (ANT, Enschende, The Netherlands), bandpass filtered between 0.01 and 170 Hz, and saved digitally using ASAlab software (ANT). The identity of each audio tone (standard/oddball) was communicated to the EEG amplifier via TTL codes sent from PsychoPy via the parallel port (Peirce, 2007).

4.1.5 Artifact Correction and Data Processing

The MNE-Python library (Gramfort et al., 2013, 2014) was used for all data preprocessing. For ERP and statistical analysis, a 0.1 to 40 Hz bandpass filter was applied to the data, followed by manual identification and removal of electrodes and epochs with excessive noise. The data were then segmented into epochs spanning 200 ms prior to the onset of each auditory tone, to 1 s after. Independent Components Analysis was then used to identify and remove artifacts such as eye blinks and movements (Delorme & Makeig, 2004) using the FastICA algorithm (Hyvarinen, 1999). EEG data was referenced to the mastoids using the average of signals from electrodes located at TP9 and TP10. The epochs that occur in the 10 s before the reported mind wandering (excluding the 1 s window before the report) were assigned a mind wandering label, while epochs that occur in the 10 s after the reported mind wandering (excluding the 1 s window after the report) were assigned an on-task label. Oscillatory analysis was performed on longer 10 s epochs which were similarly bandpass filtered at 0.1 to 40 Hz and labeled but not filtered using ICA.

4.1.6 Data Analysis

There was a high degree of variability between individuals in the number of mind wandering events reported (range: 0-61). Following best practices (Luck, 2014), we opted to control for radical inter-subject variability. Participants were excluded if they yielded 1 or fewer mind wandering responses. Three participants' data were excluded for this reason and three were excluded due to technical issues in their recording. Each participant included yielded between 12 and 140 relevant events (oddball audio tones) within the 10 s window before and after self-reported mind wandering events. This resulted in a total of 5754 epochs (2201 mind wandering standard, 466 mind wandering oddball, 2552 on task standard, 535 on task oddball).

After assessing the grand average waveform, we selected four time intervals for statistical analysis. Intervals were chosen based on observed positive and negative components, which corresponded to the P1, N1, P2 and P3. Analysis was performed at the 125 to 175 ms, 175-225 ms, 225-275 ms and 275-375 ms intervals in a region of interest centred around a posterior region centred around the Pz electrode (including electrodes Pz, Cpz, POz, CP3, CP4, P3, P4). These locations were chosen arbitrarily for exploratory analysis by visually inspecting the components. Statistical analysis was performed using linear mixed effects analysis (Tremblay & Newman, 2015). The mean amplitude over these windows were used as the dependent measure for ERP analysis. The model's fixed effects included mental state (mind wandering, on task) and stimulus type (standard, oddball); random by-subject slopes for mental state and stimulus type, as well as random intercepts for each subject, were included as well. These model parameters were selected based on comparison of Akike Information Coefficient (AIC) values (Wagenmakers & Farrell, 2004). Posthoc comparisons between the calculated LME amplitudes were performed by observing differences among the amplitudes triggered by the oddball and standard stimuli for each condition.

Oscillatory activity was compared by calculating power spectral density (PSD) of delta (1-4 Hz), theta (4-7 Hz), alpha (8-12 Hz) and beta (13-30 Hz) using multitaper method on data from the aforementioned region of interest (Percival & Walden, 1993). Average PSD values each individual were calculated and compared using parametric t-tests at both the group and individual level. Comparisons were made between pre and post mind wandering reports (10 s) as well as between the four conditions of short epochs (1.2 s).

4.1.7 Decoding

We performed decoding using two methods. Classification on the 1.2 s ERP epochs was performed on epochs which were processed using ICA, while classification on the longer 10 s epochs was performed on raw data after transforming the data using common spatial patterns. Data from participants was analyzed at the participant level to control for confounding differences between participants. The outcome of classification is a decision as to whether an epoch is from a reported mind wandering state (i.e. 10 seconds before the button press) or from the on-task state (i.e. 10 seconds after the button press). We applied five machine learning techniques to this task: linear kernel Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Logistic Regression (LR), Gaussian Naive Bayes (NB) and Random Forest (RF). Classification results were evaluated using shuffled 3-fold cross validation. We evaluated the performance of different methods using accuracy, precision and recall. Tables D.1 and D.2 in Appendix E describe the detailed structure of the data used in both frequency-series and time-series classification tasks.

4.2 Results

4.2.1 ERP Analysis

A selection of the calculated grand average from the region of interest is illustrated in Figure 4.2. The grand average yielded four visibly distinct ERP components elicited by the stimuli. We observed a visible positive component beginning at 125 ms following the stimulus presentation, as well as a negative component at 175 ms, followed by positive components at approximately 225 ms and 275 ms. Figure 4.3 highlights average amplitudes at select times in the chosen windows. Tables 4.1, 4.2, 4.3 and 4.4 summarizes the results of linear mixed effects analysis of the models. Results from the mixed effects analysis showed significant effects between the conditions among the observed P1, N1, P2, as well as P3 components, which will be discussed in turn. Quiz scores from participants who successfully completed the task with usable EEG

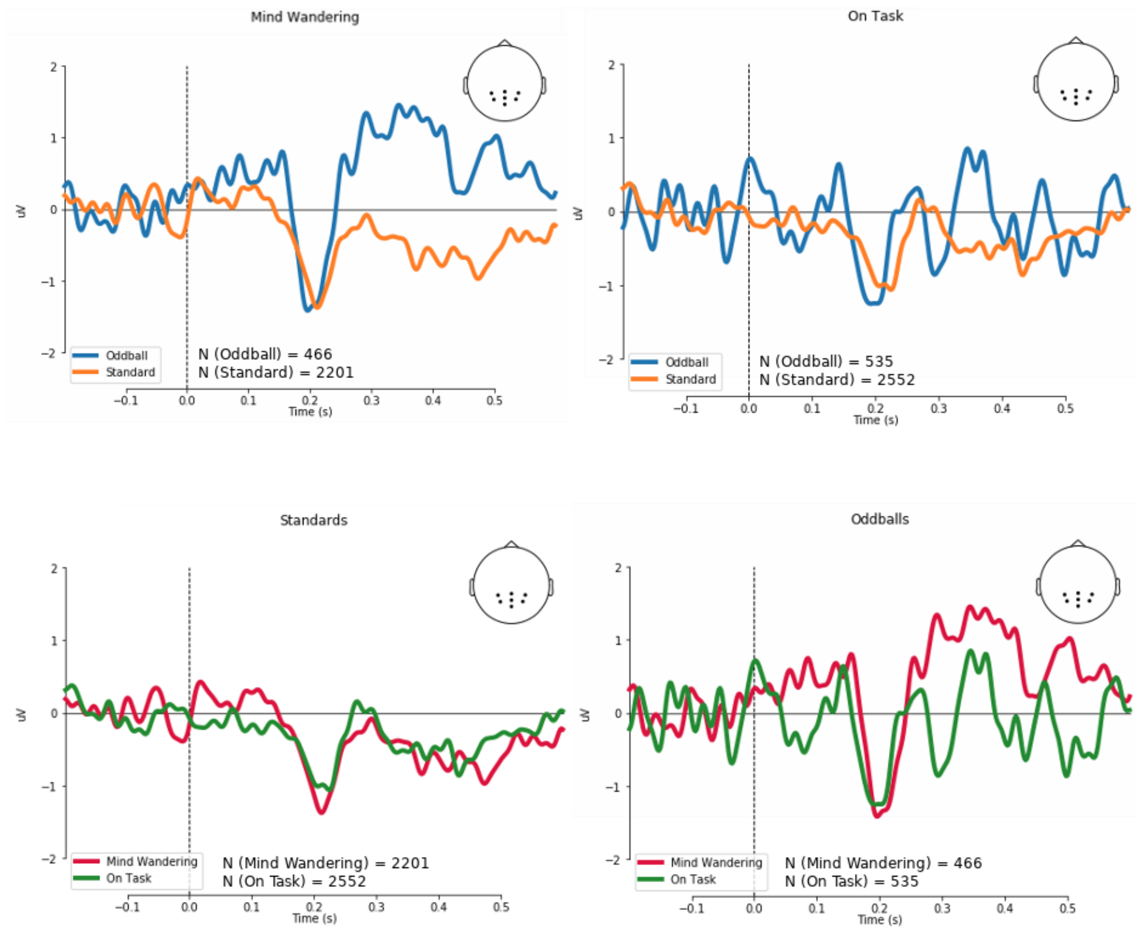


Figure 4.2: The grand average waveforms from Experiment 1 at the region of interest observed

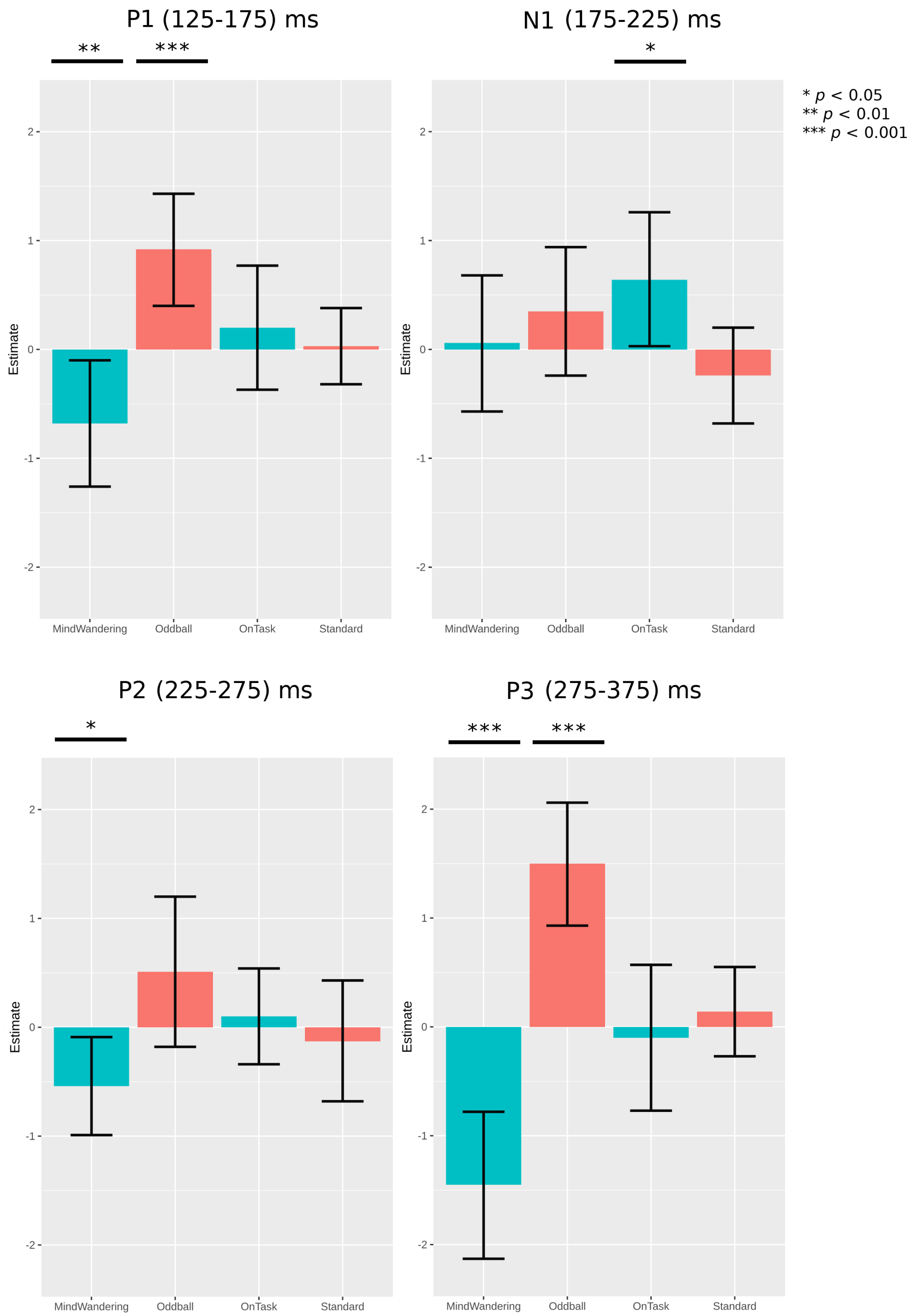


Figure 4.3: Experiment 1 component amplitudes by condition

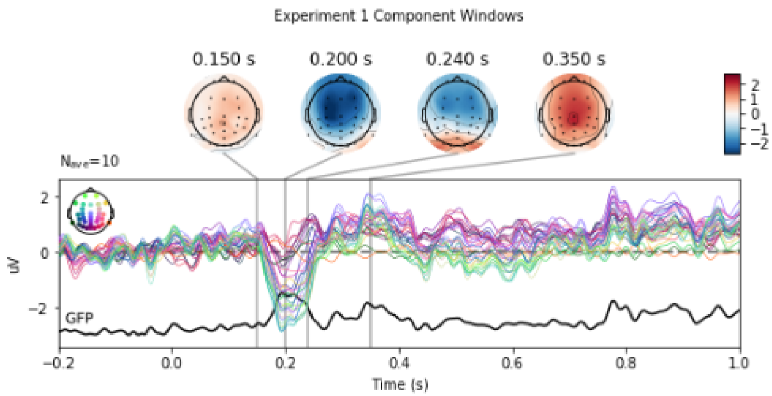


Figure 4.4: Topographic maps of components observed from Experiment 1 grand average

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	–0.68	0.230	–2.95	0.0032**
Oddball vs Standard	OT	0.20	0.227	0.89	0.7436
MW vs OT	Oddball	0.92	0.206	4.43	< 0.001***
MW vs OT	Standard	0.03	0.141	0.23	0.8182

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 4.1: P1 amplitude LME contrasts from experiment 1

were analyzed using two-tailed paired t-test. Participants were found to attain significantly lower ($p < 0.001$) pre quiz scores (mean = 1.7, SE = 1.57) as compared to post quiz scores (mean = 3.7, SE = 1.70). In both the pre and post quiz, participants correctly answered fewer than 50% of the 8 questions asked.

Highly significant effects of stimulus type during the mind wandering state ($p < 0.01$) were observed, as well as a highly significant effect of mental state among responses to oddball stimuli ($p < 0.001$) of P1 component amplitudes were observed, as

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	0.06	0.249	0.23	0.8175
Oddball vs Standard	OT	0.64	0.247	2.61	0.0364*
MW vs OT	Oddball	0.35	0.237	1.47	0.4212
MW vs OT	Standard	–0.24	0.176	–1.35	0.4212

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 4.2: N1 amplitude LME contrasts from experiment 1

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	-0.54	0.181	-3.00	0.010*
Oddball vs Standard	OT	0.10	0.175	0.580	1.000
MW vs OT	Oddball	0.51	0.276	1.86	0.190
MW vs OT	Standard	–0.13	0.223	0.574	1.000

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 4.3: P2 amplitude LME contrasts from experiment 1

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	–1.45	0.270	–5.38	< 0.001***
Oddball vs Standard	OT	–0.10	0.269	–0.36	0.785
MW vs OT	Oddball	1.50	0.226	6.62	< 0.001***
MW vs OT	Standard	0.14	0.164	0.86	0.785

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 4.4: P3 amplitude LME contrasts from experiment 1

summarized in Table 4.1. Differences in N1 responses to oddball stimuli (vs standard stimuli) were observed ($p < 0.05$) when participants reported being on task, but not when mind wandering, as summarized in Table 4.2. As with the P1 component, we observed significant differences in P2 amplitudes triggered by oddball (vs standard) stimuli when mind wandering but not when on task ($p < 0.01$), as summarized in Table 4.3.

In addition, LME analysis on P3 amplitude revealed significant main effects of both stimulus type (oddball vs standard) when participants reported being in the mind wandering state, but not when on-task ($p < 0.001$). Differences in responses to oddball stimuli between the mental states (on-task vs mind wandering) were also found to be highly significant ($p < 0.001$). Table 4.4 summarizes these findings.

4.2.2 Oscillatory Activity

Comparisons of average epoch power among the long (10 s) epochs among all individuals were not significant at $\alpha = 0.05$ and is summarized in Table 4.5. Comparisons of the short (1 s) epochs revealed significant differences in oscillatory activity elicited by stimulus type, but not among mental states, as summarized in Table 4.6. Differences in theta (4-7 Hz) activity triggered by oddball stimuli when mind wandering were

Band	Frequency (Hz)	t	p (raw)
delta	1-4	0.3172	0.7547
theta	5-7	-0.0417	0.9671
alpha	8-12	0.4729	0.6419
low beta	13-21	-0.1238	0.9028
beta	13-30	0.1524	0.8805

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 4.5: Results of comparisons of average band power spectral density among participants based on 10 second (long) epochs

Comparison	Frequency (Hz)	Condition	t	p (raw)	p (Holm)
Stimulus Type	4-7	MW	-4.0101	< 0.001***	< 0.001***
Stimulus Type	4-7	OT	-2.5482	0.0201*	0.1206
Stimulus Type	8-12	MW	-3.4668	0.0027**	0.0243*
Stimulus Type	8-12	OT	-2.1234	0.0478*	0.239
Stimulus Type	13-30	MW	-2.5784	0.0189*	0.132
Stimulus Type	13-30	OT	-2.9317	0.0089**	0.0712
Mental State	4-7	Standard	0.7139	0.4844	0.4845
Mental State	4-7	Oddball	0.3137	0.7572	0.757
Mental State	8-12	Standard	0.5059	0.6190	0.6190
Mental State	8-12	Oddball	0.5854	0.5655	0.5655
Mental State	13-30	Standard	0.6195	0.5433	0.5433
Mental State	13-30	Oddball	0.0302	0.9761	0.9761

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 4.6: Results of comparisons of average band power spectral density among participants based on 1 second (short) epochs

Part. No.	No. Trials	p (mean theta)	p (mean alpha)	p (mean beta)
3	74	0.9792	0.0034**	0.0001**
4	60	0.1589	0.0033*	0.0463*
6	16	0.9120	0.0415*	0.7732
7	60	0.3753	0.9059	0.0295*
8	12	0.6599	0.7347	0.5906
9	122	0.6051	0.0359**	0.0963
11	52	0.4111	0.7932	0.5854
12	14	0.8977	0.2066	0.3358
13	40	0.1418	0.3966	0.0016**
14	24	0.2206	0.0268*	0.0253*

* Significant at $\alpha = 0.05$); ** Significant at $\alpha = 0.01$

Table 4.7: Comparisons of frequencies band average power over 10 second epochs between conditions for each participant

highly suggestive. Power comparisons were also conducted on long epochs for each individual, revealing significant differences in alpha (8-12 Hz) and beta (13-30 Hz) frequency band power during the mind wandering and on task states among some participants, though this analysis runs the risk of a Type I error due to the number of comparisons made. The results of the individual oscillatory analysis is summarized in Table 4.7.

4.2.3 Classification Performance

We summarize classification accuracy of various machine learning algorithms and tasks in Table 4.10. We were unable to create a time series classifier that was able to classify more accurately than random chance. We were able to create a number of frequency domain classifiers using Common Spectral Patterns and a Linear Discriminant Analysis classifier which performed better than random chance, including a classifier that performed with a mean of 73.7 % accuracy on the 10 participants. We also explored other classifiers that used the CSP preparation technique which produced similar classification results which were not statistically different from this classifier in classification accuracy.

Participant Number	Pre	Post
3	37	37
4	30	30
6	8	8
7	30	30
8	6	6
9	61	61
11	26	26
12	7	7
13	20	20
14	12	12
Total	237	237

Table 4.8: Description of frequency domain classes Experiment 1 decoding analysis

Participant Number	Pre Oddball	Pre Standard	Post Oddball	Post Standard
3	76	403	79	338
4	72	329	65	296
6	16	96	23	88
7	67	330	60	294
8	11	73	23	60
9	150	656	102	490
11	57	247	40	208
12	14	79	14	83
13	44	213	34	205
14	28	126	26	139
Total	535	2552	466	2201

Table 4.9: Description of time domain classes for Experiment 1 decoding analysis

Participant	LDA 7-13 Hz	LDA 1-30 Hz	LDA 4-30 Hz	SVM 4-30 Hz	NB 4-30 Hz	RF 4-30 Hz
3	0.716	0.838	0.689	0.689	0.703	0.703
4	0.617	0.633	0.483	0.717	0.600	0.617
6	0.625	0.750	0.563	0.750	0.75	0.750
7	0.683	0.650	0.683	0.717	0.667	0.783
8	0.500	0.667	0.500	0.750	0.667	0.583
9	0.713	0.713	0.672	0.730	0.672	0.713
11	0.711	0.788	0.712	0.712	0.538	0.731
12	0.500	0.786	0.500	0.500	0.426	0.357
13	0.875	0.875	0.825	0.875	0.875	0.875
14	0.792	0.667	0.500	0.541	0.583	0.583
Mean	0.673**	0.737***	0.613*	0.698**	0.648**	0.670**

* Significantly better than chance ** Insignificant from best classifier at $\alpha = 0.05$
*** Best classifier

Table 4.10: Comparison of classification accuracy from select classifiers. Mean accuracy is provided for reference

4.3 Discussion

The primary motivation of this study was to replicate the findings of Braboszcz and Delorme (2011) in an e-learning setting, specifically that of an online lecture. We did not succeed at replicating these results and instead discovered evidence for a different effect. Though we observed significant differences in each of the P1, N1, P2 components, these were centred in the parietal region, unlike the frontal region described by Braboszcz and Delorme (2011). Similar to their study, P1 and P2 components were significantly more positive when in a reported mind wandering state, while N1 components were significantly more negative when reporting being on task. However, responses to P3 stimuli were significantly more positive when in either mind wandering and on-task states, which was not described by Braboszcz and Delorme (2011). Unlike the P1-N1-P2 complex, the P3 component is not obligatory and is typically only present to task relevant stimuli. This suggests that participants' conscious attention was consistently sifted towards the auditory stimuli immediately before and after a mind wandering report.

These observations were consistent with our hypotheses H1a and H1c and H1d, which posited that we would observe differences between the sates and observe greater

P1 and P2 amplitude, but also a P3 effect. The results also lend evidence for the converse of hypothesis H1b. We believe that these are the result of user attention is being re-focused on the oddball stimuli during a mind wandering state, and away from the task of learning. It is possible that the P3 is observed in both mind wandering and on-task states because of the nature of the button press report. Participants become aware of the sounds when mind wandering. They then continue to attend to the sounds as they complete the task of a button press.

We did not observe the same oscillatory activity as described by Braboszcz and Delorme (2011). They employed time-frequency map decompositions to determine differences in oscillatory activity between the mind wandering and on-task states and discovered significant results. We did not observe any relationships between mental states at the aggregate level. However, we did observe differences in alpha and beta band activity at the individual level. This possibly explains how the individual decoding task using oscillatory data exhibited such strong classification results. It is likely that there are significant differences in band activity that are not observable in the aggregate but are occasionally distinct at the individual level. A larger sample may observe statistical differences of this small effect size. These results lend strong support for H5, which posited that we could construct an algorithm which could measure differences in reported mind wandering states.

The best performing classifier used the CSP pre-preparation process on frequency ranges 1 Hz to 30 Hz with a LDA classifier. One problem with this decoder is that it is dominated by delta activity. Delta frequencies are also indicative of blinks, which suggests that this may not be the product of brain data. The attentional blink is a well-studied phenomenon that is associated with attentional stimuli, and could have been triggered by auditory stimuli (Horváth & Burgyán, 2011). Such activity could potentially have an amplification effect on the classification algorithm, as it detects delta activity that is actually the product of such blinks. Other classifiers which included CSP with frequencies 4 Hz to 30 Hz (theta, alpha beta) produced statistically similar results however, which is why they are described in Table 4.9. The SVM classifier managed to produce a particularly strong result, regularly classifying with over 70% accuracy.

4.3.1 Limitations

This experiment was exploratory in nature, and was designed to replicate previously-reported results in a novel setting. We found statistically significant differences which corroborated parts of our first hypothesis, but did not lend enough evidence to be completely confident. Notably, the timing of the potentials and consequently our contrasts (125-175 ms for P1, 175-225 ms for N1, 225-275 ms for P2, 275-375 ms for P3) for were very late and were different from those described by Braboszcz and Delorme (2011), who reported a P2 from 180-280 ms. The differences in timing is not surprising because of the differences in task in paradigm. However, upon investigating our experimental setup we discovered a delay in the TTL recording and the timing recorded by the experiment computer. The delay caused a gap between the two timings, which resulted in a delayed ERP response. Later experiments on this gap detected a consistent delay on between our computing system and audio record time (mean 51.78 ms, SD = 10.345 ms), though the experiment results could have been further delayed to an even greater degree during the time of recording. This phenomenon thus warrants continued investigation. Successful replication of the results in a similar setting with the recording error corrected would add considerable evidence to the usefulness of the event-related potentials being observed, while also eliminating concerns about sample sizes, technical errors, statistical and classification analysis.

A second unrelated related concern is that it is not clear that the phenomenon being observed truly represents mind wandering; it could be an effect of the button press report method. Though participants were given instructions about what constitutes mind wandering and how to observe it, there is a high degree of subjectivity inherent in the task of self-reporting the timing. The high variance in button press responses (ranging from 6 to 61) suggests that a different experience sampling method could be used to determine correlates. It is possible that the observed response is unique to this e-learning context or specific video. In addition, previously reported psychometric scales could be used to determine whether the ERP and oscillatory activity observed is correlated with the mind wandering phenomenon. A replication study could employ such techniques to determine correlates.

A potential issue with our classification analysis is that some classification tasks had very small sample sizes. Though we utilized 3-fold cross validation to mitigate the impacts of small datasets, this technique is prone to overfitting the data. In addition, the classification analysis was performed offline. Though decoding was performed on minimally filtered data, it nonetheless used a band pass filter, which may limit its potential applications to real-time technology such as brain-computer intferface (BCI)s.

We were thus motivated to conduct a second study with a similar design but with a number of small differences. The second study should manage the TTL recording error, but replicate the ERP results earlier than the components described in this study. The second study was envisioned to involve a different experience sampling measure such as that described by Wammes and Smilek (2017) to help ensure that the phenomenon being observed accurately describes mind wandering. Furthermore, the study should include a larger sample size, a post-hoc measure and a measure of rote learning. Successful replication of the key findings in this paper would lend considerable evidence that we are able to observe mind wandering by observing differences in the P1, N1, P2, and P3 components, as well as neural oscillations. A disadvantage of designing a study using such an experience sampling method is that it would not be able to yield sufficient reports for performing the CSP decoding analysis described in this study.

4.4 Summary

In this chapter we described an experiment to replicate the results of a prior experiment by Braboszcz and Delorme (2011), but in an online lecture setting. We observed many of the phenomena that they described, including an amplified P2. However, the observed differences were greater in the parietal region, rather than frontal, and we observed different oscillatory activity. Importantly, we observed a distinct P3 effect, rather than just a P2 effect. This experiment provides evidence that we are able to measure mind wandering using neurophysiological indicators during an online lecture session, and provides evidence that we can measure it in real time using electroencephalography (EEG) decoding. In the following chapter, we describe an experiment to replicate these findings at a larger scale and corroborate the hypotheses not addressed by this study.

Chapter 5

An Experiment to Measure the Effect of Mind Wandering on E-Learning Efficacy

Following the experiment described in Chapter 4, we are left with some questions about the neurophysiology of mind wandering. Though we successfully identified a relationship between amplitudes and mind wandering experience samples, the relationship observed featured a distinct P3 component, normally exhibited by the onset of task-related stimuli. Furthermore, we did not observe the oscillatory activity reported by Braboszcz and Delorme (2011). It remains to be seen whether this is the result of the neurophysiology of mind wandering, rather than a product of the sampling technique employed.

In this chapter we describe the results of a second study that sought to replicate many of the key findings described in Chapter 4, and extend its conclusions. We incorporated a number of small changes to the experiment design. We drew from a larger sample size, incorporated a different experience sampling method and employed technology that overcame the technical recording error previously discovered. In addition, we employ some of the ex post measures described by Sullivan et al. (2015), the NASA Task Load Index (TLX) (1988), and a multiple choice quiz to measure rote learning. In the case of the multiple choice quizzes, we administered the quiz before the video and again following the video in an effort to control for task difficulty. The objective of this study was to uncover evidence that the phenomenon being observed reflects mind wandering, and to examine mind wandering's impact on rote learning.

5.1 Methods

5.1.1 Participants

52 healthy students (36 females 16 males; aged 17-28 years; mean 20.6 and SD 2.5) gave written consent to participate in the experiment. Participants were excluded

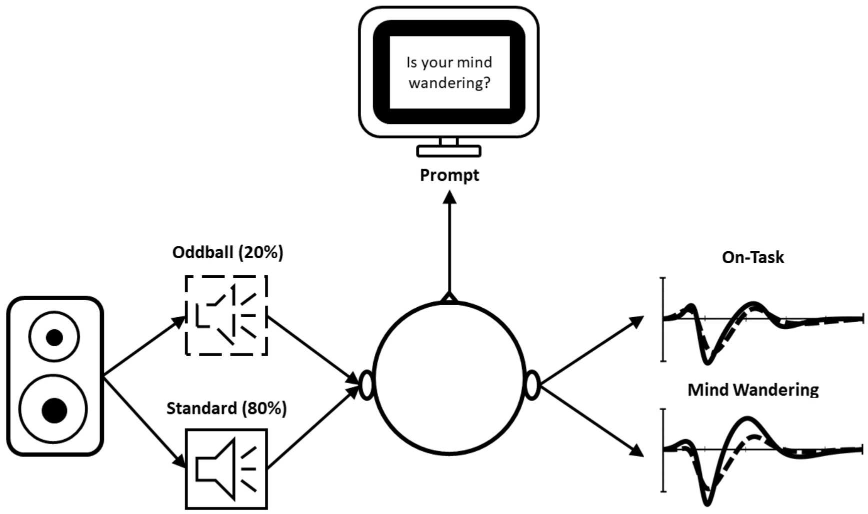


Figure 5.1: An illustration of the experiment design used in this experiment

if they were not fluent in English, identified as having neurophysiological disorders or were taking medication that could lead to abnormal EEG. Participants were also excluded if they had taken a course in venture capital, the subject of the video. Participants provided written and informed consent and were financially compensated CAD $25 for their time. All procedures were reviewed by Dalhousie University's Research Ethics Board.

5.1.2 Experiment Stimuli

Audio stimuli consisted of 100 ms tones; the frequent standard tones were 500 Hz and the oddball tones were 1000 Hz. Oddballs and standard conditions were assigned to each tone randomly at a 80:20 ratio.

5.1.3 Procedure

After completing the informed consent procedure, participants were again fitted with the EEG cap and brought to the testing room. Participants were asked to attend to a 75 minute English language video on the subject of venture capital (Fu, 2017). The subject matter and video were chosen because it had some utility to the participants and was on a subject not commonly taught to our subject population. Pilot testing suggested that this video would trigger variations in mind wandering and attention for most participants. The video consisted exclusively of two lecturers talking as well as questions from the lecture hall audience. Study participants were asked to pay attention to the video and ignore the audio stimuli, which were presented every 1-1.5 seconds (mean 1.25). 10 mind wandering prompts were triggered pseudo-randomly throughout the video. At each prompt, participants were asked to report their degree of mind wandering or on-task experience from the time period immediately before the mind wandering prompt (Wammes & Smilek, 2017). The timings of the prompts are summarized in Table D.4 in Appendix E. The PsychoPy library was used to present the audio stimuli and record manual responses (Peirce, 2007).

5.1.4 EEG Recording

Participants were fitted with 32-channel scalp electrodes (ActiCap, BrainProducts GmbH, Munich, Germany) positioned at standard locations according to the International 10-10 system and referenced during recording to the midline frontal (FCz) location. Bipolar recordings were made between the outer canthi of the two eyes and above and below one eye, to monitor for eye movements and blinks. Electrode impedances were kept below 15 kOhm throughout the experiment. Electroencephalography data were sampled at 512 Hz using Refa8 amplifier (ANT, Enschende, The Netherlands), and bandpass filtered between 0.01 and 170 Hz, and saved digitally using the ASAlab software (ANT). The identity of each audio tone (standard/oddball) was communicated to the EEG amplifier via TTL codes sent from PsychoPy via the parallel port (Peirce, 2007). To compensate for the difference in timing between the computer and TTL port, an Arduino device was configured to record the precise moment of the audio sound using the TTL port (Baker, 2013).

5.1.5 Artifact Correction and Data Processing

The MNE-Python library (Gramfort et al., 2013, 2014) was used for all data preprocessing. All audio events were adjusted by calculating the differences between the Arduino trigger and the computer recorded timings and adjusting the epoch times using MNE Python. For ERP and statistical analysis, a 0.1 to 40 Hz bandpass filter is applied to the data, followed by manual identification and removal of electrodes and epochs with excessive noise. The data were then segmented into epochs spanning 200 ms prior to the onset of each auditory tone, to 1 s after. Independent components analysis was then used to identify and remove artifacts such as eye blinks and movements (Delorme & Makeig, 2004) using the FastICA algorithm (Hyvarinen, 1999) and referenced to the mastoids by averaging the TP9 and TP10 electrodes. The epochs analyzed were extracted from the 10 seconds before a mind wandering prompt and labeled based on user responses to the prompts (see Appendix D). Participant responses were labeled as either "on task" or "mind wandering" based no their responses, while "both on task and mind wandering" responses were assigned to the minority class. Oscillatory analysis was performed on longer 10 s epochs which were similarly bandpass filtered at 0.1 to 40 Hz and labeled using the protocol above, but not filtered using ICA. Power spectrum density was calculated at the theta (4-7 Hz), alpha (8-12 Hz) and beta (13-30 Hz) frequency bands using the multitaper method (Percival & Walden, 1993). The trigger delay identified in Chapter 4 was corrected by basing the event related potentials around the audio timings recorded by the Arduino device rather than the timings recorded by the computer.

5.1.6 Data Analysis

Given that there were exactly 10 mind-wandering prompts for each participant, there was no variability in the number of responses, though there was a variability in the degree of mind wandering reported. All participants were included in the analysis if they reported some degree of both mind wandering or on-task experience (i.e. selected being either "completely on-task" or "somewhat on task" or "neither on task nor mind wandering" in addition to some degree of mind wandering). Three participants' data were excluded for lack of variance in responses while and five were excluded due to technical issues in their recording. Data was prepared in a way to contrast between on task and mind wandering states according to a simple protocol. All observations observed in the 10 seconds before a "completely on task" or "somewhat on task" report were labeled as "on task", and conversely for the "mind wandering" label. Instances preceding a "neither on task nor mind wandering" label were assigned to either "on task" or "mind wandering" depending on which task was the minority state. The result is two labels which can be used to contrast between higher degrees of perceived mind wandering versus lower degrees for each individual. Each participant included yielded between 1 and 28 oddball audio tones within the 10 s window before the mind wandering prompt. This resulted in a total of 6552 epochs (2303 mind wandering standard, 440 mind wandering oddball, 3201 on task standard, 608 on task oddball).

After assessing the grand average waveform, we selected four time intervals for statistical analysis. Intervals were initially chosen by subtracting the observed mean difference (52 ms) between the Arduino signals and the computer recorded signals from the previously identified component times. However, these were observed to not conform to the expected components, as illustrated by figure D1 in Appendix E. We noticed that components were approximately 100 ms earlier than in Experiment 1, so analysis was performed at the 25-75 ms, 75-125 ms, 125-175 ms and 175-275 ms intervals in a new thought region of interest centred around the Pz electrode (including electrodes Pz, Cpz, POz, CP3, CP4, P3, P4). Statistical analysis was performed using linear mixed effects analysis and generalized additive models (Tremblay & Newman, 2015). The mean amplitude over these windows were used as the dependent measure for ERP analysis. Similarly, oscillatory theta, alpha and beta power spectrum density was calculated using the multitaper method and used as the dependent measure. The model's fixed effects included mental state (mind wandering, on task) and stimulus type (standard, oddball); random by-subject slopes for mental state and stimulus type, as well as random intercepts for each subject, were included as well. These model parameters were selected based on comparison of Akike Information Coefficient (AIC) values (Wagenmakers & Farrell, 2004). Due to the small number of mind wandering reports, decoding analysis was not performed on this data.

5.2 Results

5.2.1 EEG Correlates of Mind Wandering and On Task States

A selection of the calculated grand average results at the region of interest is illustrated in Figure 5.2. The grand average resulted in three visibly distinct ERP components that are elicited by the stimuli. As expected, the components were visible earlier than in the first experiment due to the audio timing correction, though they were earlier than the 52 ms expected, as depicted in Figure 5.4 and 5.5. We observed a visible positive component beginning at 25 ms following the stimulus presentation, as well

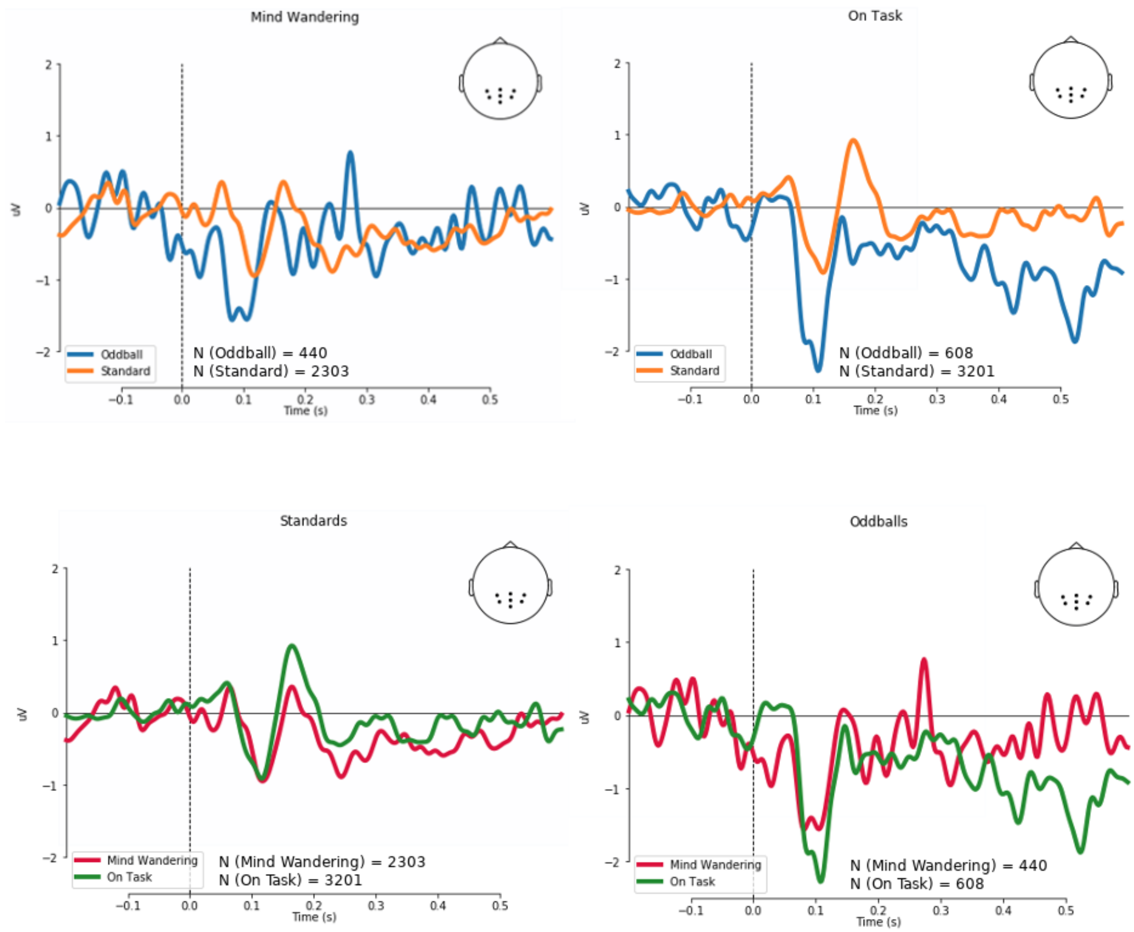


Figure 5.2: The grand average waveforms from Experiment 2 at the region of interest observed

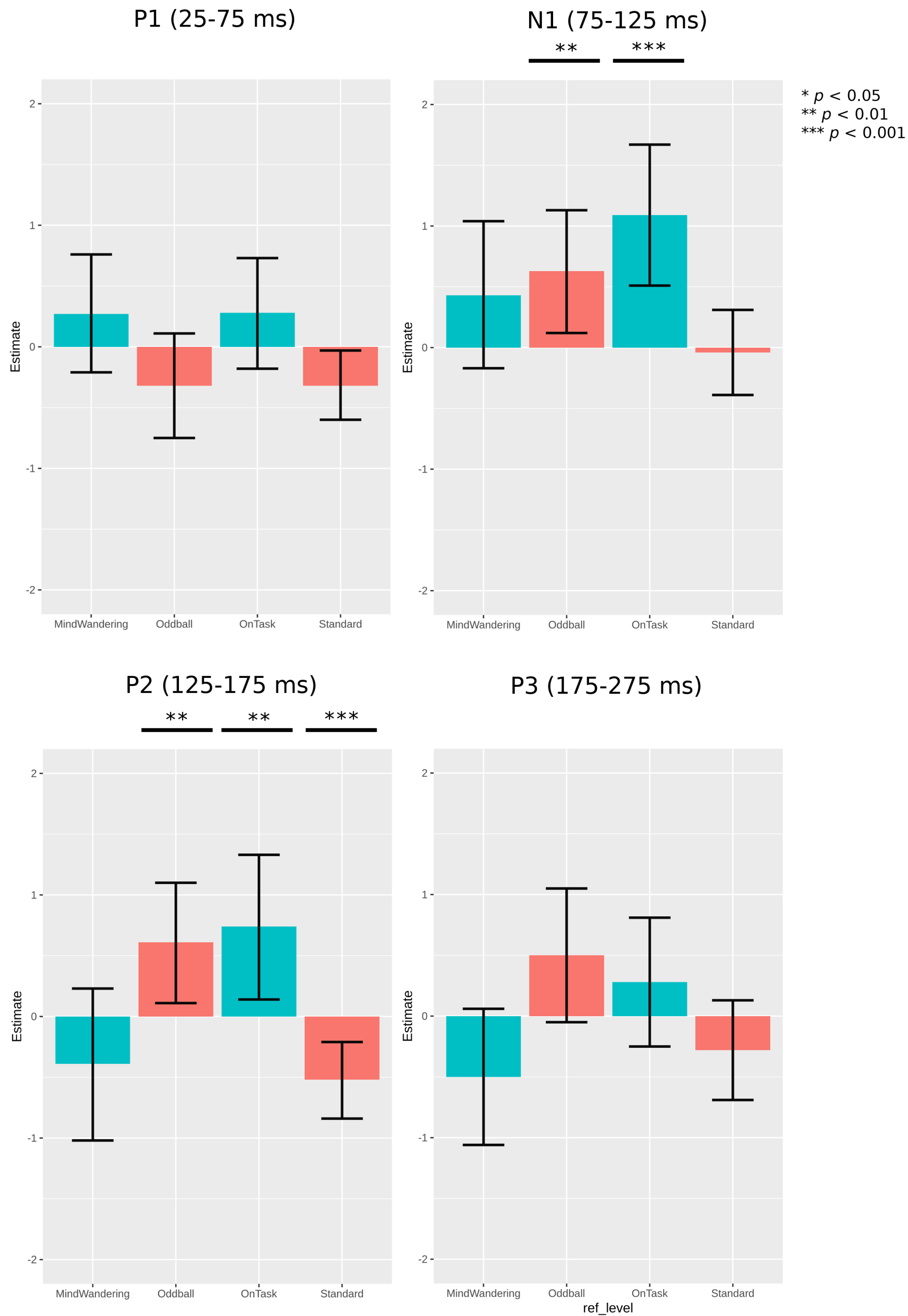


Figure 5.3: Component amplitudes by condition

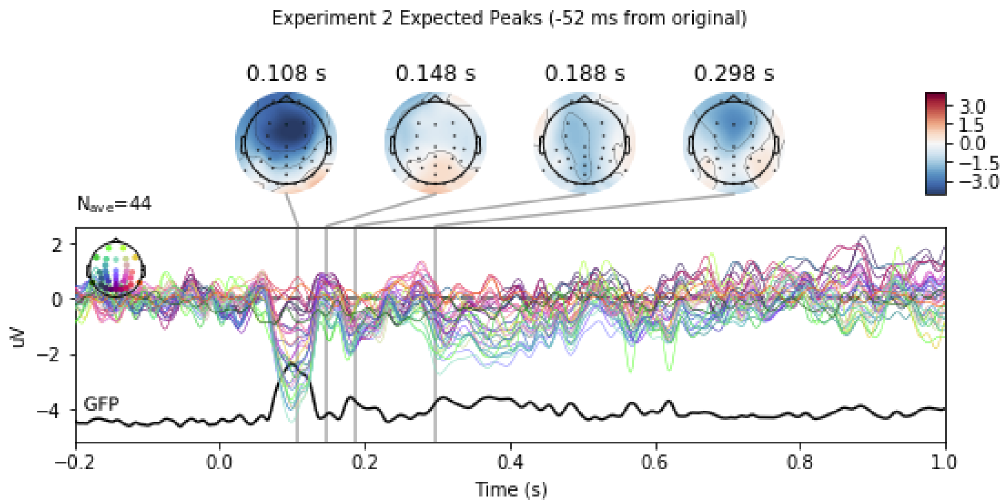


Figure 5.4: Expected component peaks from Experiment 2, assuming the mean timing differences were constant

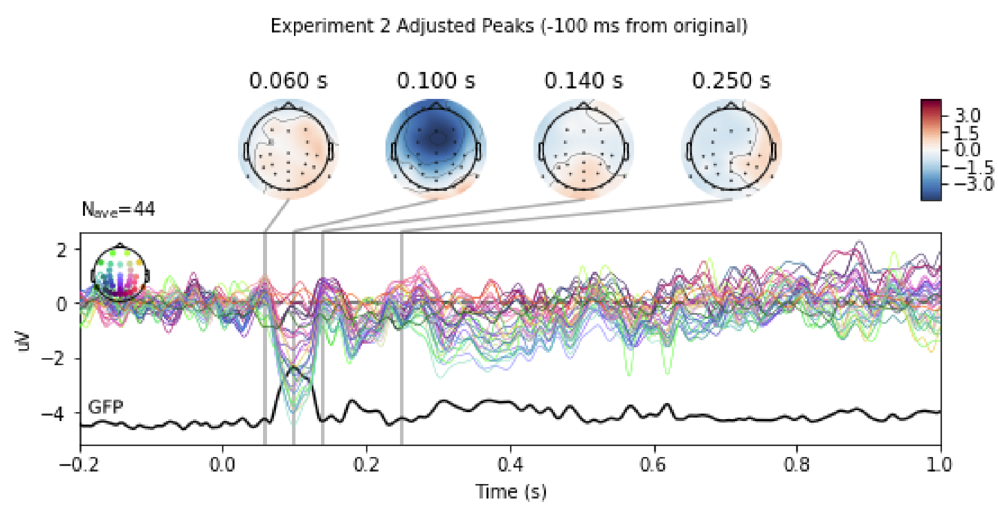


Figure 5.5: Topographic maps of components observed from Experiment 2 grand average

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	0.27	0.193	1.41	0.262
Oddball vs Standard	OT	0.28	0.183	1.51	0.262
MW vs OT	Oddball	–0.32	0.172	–1.84	0.196
MW vs OT	Standard	–0.32	0.114	–2.79	0.021*

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 5.1: P1 amplitude LME contrasts from experiment 2

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	0.43	0.242	1.79	0.148
Oddball vs Standard	OT	1.09	0.232	4.72	< 0.001**
MW vs OT	Oddball	0.63	0.201	3.12	0.006**
MW vs OT	Standard	–0.04	0.139	–0.26	0.793

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 5.2: N1 amplitude LME contrasts from experiment 2

as a negative component at 75 ms, followed by a positive component beginning at 125 ms. These correspond to a regular P1-N1-P2, as before, but were earlier than expected. The P3 component was not visible but was nonetheless examined.

Figure 5.3 illustrates component amplitude averages by condition at the chosen windows. Tables 5.1, 5.2, 5.3 and 5.4 summarize the results of linear mixed effects analysis of component amplitudes. No significant main or interaction effects were observed among P1 amplitudes. Significant differences in response to stimuli (standard vs oddball) were observed when participants reported being on task ($p < 0.001$) but not when they reported mind wandering. In addition, significant differences in responses to oddball stimuli among states (mind wandering vs on task) were observed ($p < 0.01$).

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	–0.39	0.250	–1.58	0.115
Oddball vs Standard	OT	0.74	0.239	3.08	0.006**
MW vs OT	Oddball	0.61	0.198	3.08	0.006**
MW vs OT	Standard	–0.52	0.125	–4.16	< 0.001***

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 5.3: P2 amplitude LME contrasts from experiment 2

Contrast	Reference	Estimate	SE	t value	p-value (Holm)
Oddball vs Standard	MW	–0.50	0.224	–2.25	0.090
Oddball vs Standard	OT	0.28	0.212	1.32	0.188
MW vs OT	Oddball	0.50	0.220	2.28	0.090
MW vs OT	Standard	–0.28	0.164	–1.72	0.172

* Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 5.4: P3 amplitude LME contrasts from experiment 2

Band	Frequency (Hz)	t	p-value
delta	1-4	0.8497	0.3978
theta	5-7	-0.1316	0.8955
alpha	8-12	0.2775	0.7820
beta	13-30	0.0714	0.9431

* Significant at $\alpha = 0.05$); ** Significant at $\alpha = 0.01$

Table 5.5: Results of comparisons of average band power among participants based on 10 second (long) epochs from Experiment 2

We also observed significant effects at the P2 component. P2 responses to oddball stimuli were different among the mental states ($p < 0.01$), as were responses to standard stimuli ($p < 0.001$). In addition, responses to stimulus type (oddball vs standard) were different when on task ($p < 0.01$), but not when mind wandering. No significant effects were observed on P3 amplitude. Furthermore, we did not discover significant differences in oscillatory activity between mind wandering and on-task states, as summarized in Tables 5.5 and 5.6. Though we observed some significant differences at the individual level, these are likely circumstantial and the product of Type I errors due to the umber of comparisons made.

5.2.2 Correlates of Real-Time and Ex-Post Measures

Descriptive results of experience samples collected are provided in Table 5.9. In line with Wammes and Smilek (2017), we observed increased degrees of mind wandering as the lecture progressed ($p < 0.001$), noticing a pronounced difference between samples collected at the 15 minute and 30 minute marks. We did not observe significant relationships between the average oddball amplitudes and any of the ex post measures, as summarized in Table 5.11.

Participant	No. Trials	p (mean theta)	p (mean alpha)	p (mean beta)
1	10	0.0159*	0.8017	0.5801
5	10	0.0408*	0.5723	0.8596
14	10	0.4358	0.0143*	0.3362
27	10	0.3030	0.0016*	0.0900
29	10	0.0411	0.0034*	0.1676
31	10	0.0292*	0.9041	0.49
34	10	0.4463	0.5350	0.0281*
39	10	0.0385*	0.1798	0.0659

* Significant at $\alpha = 0.05$); ** Significant at $\alpha = 0.05$ (Holm corrected)

Table 5.6: Participants from Experiment 2 with at least one significant (raw) difference in frequency bands between the reported states

Prompt	Elapsed Time (approx.)	Mean Response
1	4 min, 35 sec	2.18
2	8 min, 32 sec	1.84
3	10 min, 20 sec	1.84
4	15 min, 8 sec	1.95
5	30 min, 8 sec	3.34
6	42 min, 50 sec	3.00
7	53 min, 15 sec	3.23
8	56 min, 15 sec	3.05
9	68 min, 7 sec	3.02
10	71 min, 7 sec	2.55

Table 5.7: Summary of mind wandering prompt timings and mean response scores

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	2.00	0.094	21.197	< 0.001***	
Time (minutes)	0.016	0.002	7.541 < 0.001***	0.112	

* Significant at $\alpha = 0.5$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 5.8: Summary of linear regression of experience sample responses and time (minutes)

P1 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-squared
(Intercept)	0.8297	0.5114	1.62	0.11	
Tech. rel. MW	-0.0414	0.0323	-1.28	0.21	
Tech. unr. MW	-0.0061	0.0235	-0.26	0.80	
NASA TLX	-0.0528	0.0366	-1.44	0.16	
Whole Model	-	-	1.54	0.219	0.0362

N1 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-squared
(Intercept)	-0.7327	0.7527	-0.97	0.34	
Tech. rel. MW	0.0552	0.0477	1.16	0.25	
Tech. unr. MW	-0.0507	0.0347	-1.46	0.15	
NASA TLX	-0.0459	0.0540	-0.85	0.40	
Whole Model	-	-	1.38	0.262	0.026

P2 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-squared
(Intercept)	-0.2466	0.7235	-0.34	0.74	
Tech. rel. MW	0.0512	0.0458	1.12	0.27	
Tech. unr. MW	-0.0376	0.033	-1.13	0.27	
NASA TLX	0.0106	0.0519	0.21	0.84	
Whole Model	-	-	0.61	0.613	-0.028

P3 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-squared
(Intercept)	-0.3315	0.7215	-0.46	0.65	
Tech. rel. MW	0.0461	0.0457	1.01	0.32	
Tech. unr. MW	-0.0252	0.0332	-0.76	0.45	
NASA TLX	-0.0001	0.0251	0.00	1.00	
Whole Model	-	-	0.408	0.748	-0.043

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.9: Summary of best fit multivariate linear regression models of experience sample means and average oddball amplitudes

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	3.483e+01	0.617e+00	21.543	$< 0.001^{***}$	
Beta	1.767e+10	9.357e+09	1.888	0.0659†	0.0563

†Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.5$; ** Significant at $\alpha = 0.01$;

Table 5.10: Summary of post-hoc linear regression of NASA TLX and beta band activity

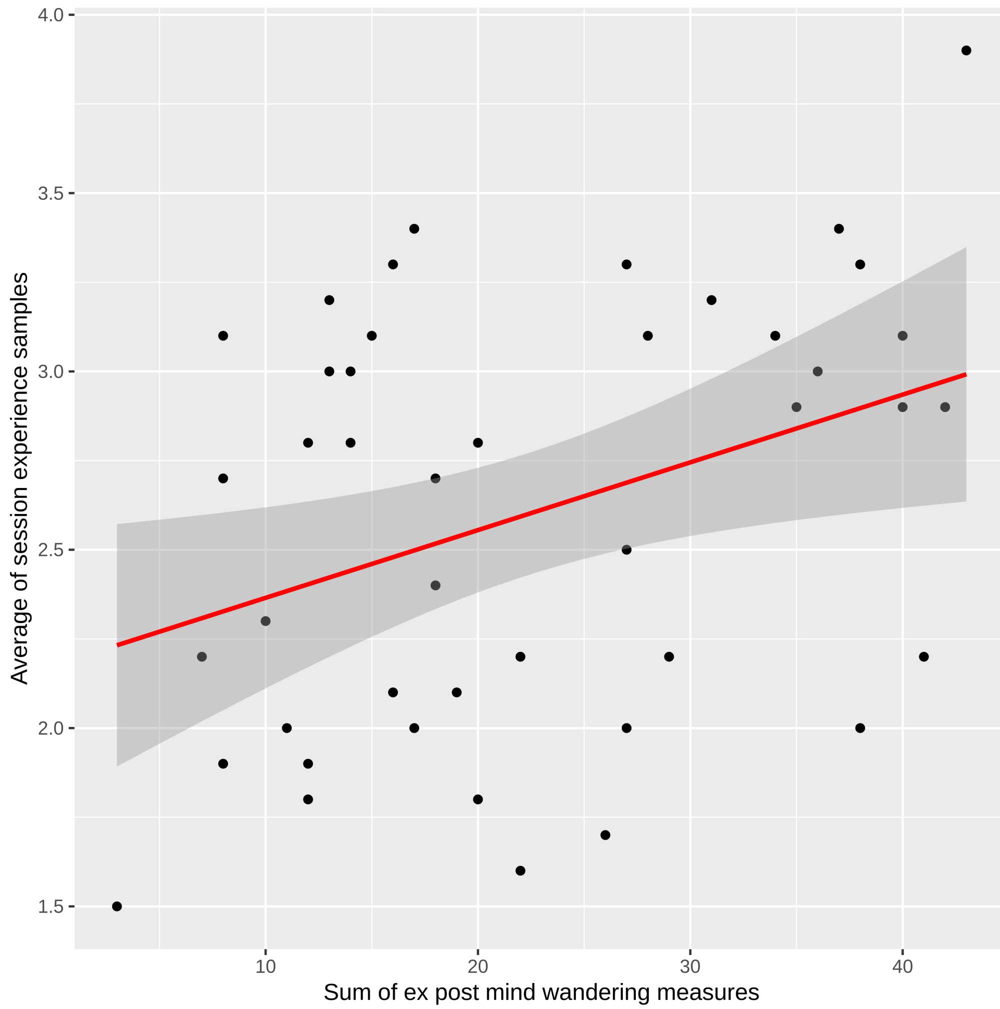


Figure 5.6: Summary of multivariate linear regression model of ex post mind wandering measures and experience sample averages

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	2.1743	0.1871	11.62	< 0.001***	
Tech. unr. MW	0.0083	0.0114	0.73	0.469	
Tech. rel. MW	0.0364	0.0161	2.26	0.029*	
Whole Model	-	-	3.97	0.0264*	0.122

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.11: Summary of multivariate linear regression model of ex post mind wandering measures and experience sample averages

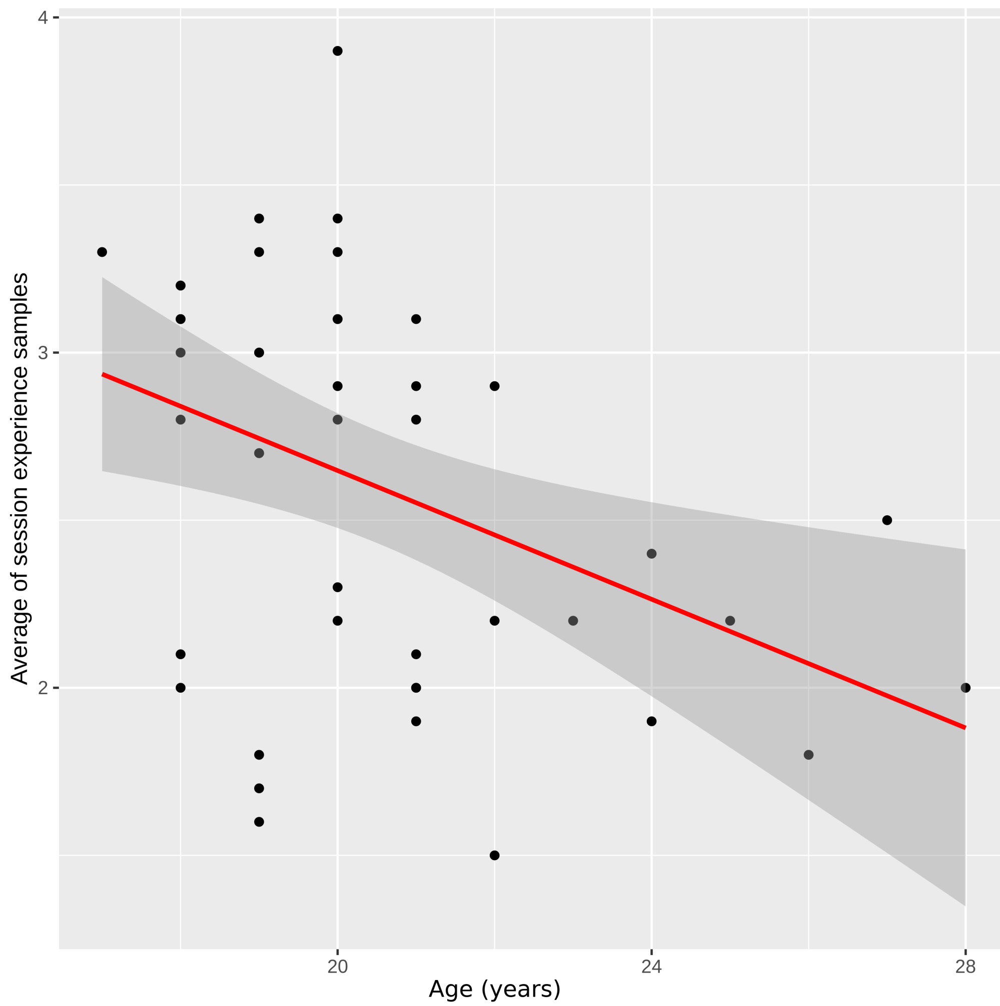


Figure 5.7: Post-hoc linear regression of mean experience sample reports and participant age.

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	3.332	0.245	13.57	< 0.001***	
Age	–0.076	0.032	–2.36	0.023*	
Gender (male)	–0.460	0.171	–2.68	0.011*	0.254

* Significant at $\alpha = 0.5$; ** Significant at $\alpha = 0.01$; *** Significant at $\alpha = 0.001$

Table 5.12: Summary of post-hoc multivariate linear regression analysis of mean experience sample reports including coefficients age and gender

We also observed a significant positive relationship between average experience sample reports and ex post mind wandering measures, as depicted in Figure 5.6 and Table 5.13. Post hoc analysis also revealed significant negative relationships between mean experience sample response and age ($p < 0.05$) and gender ($p < 0.05$). These results are summarized in Figure 5.7 and Table 5.14. Post hoc analysis also revealed an interesting relationship between beta activity ($p < 0.1$) and NASA TLX scores, summarized in Table 5.12.

5.2.3 Correlates of Rote Learning

Participants were again found to attain significantly lower ($p < 0.001$) pre quiz scores (mean = 2.86, SE = 1.27) as compared to post quiz scores (mean = 4.82, SE = 2.18), which suggests that they learned from the video. In both the pre and post quiz, participants again correctly answered fewer than 50% of the 10 questions asked. Multiple linear regression analysis did not reveal any relationship between improvement in quiz scores and ERP amplitudes, summarized in Table 5.14.

Analysis of quiz delta and survey instruments revealed a significant model using mind wandering report measures ($p < 0.01$) with a significant mean experience sample coefficient ($p < 0.05$), as depicted in Table 5.16. Univariate regression of the ex post mind wandering reports and mean experience sample scores revealed significant relationships between the various mind wandering measures and improvement in test scores, as depicted in Tables 5.17 and 5.18, 5.19 and Figure 5.8. Task load was not found to be a significant predictor of quiz performance. Linear regression of quiz deltas and NASA TLX scores are provided in Figure 5.9 and Table 5.20. This suggests that mind wandering was a significant factor in learning during this experiment while task load was not.

5.3 Discussion

These results lend considerable evidence that the N1, P2 and P3 components are useful measures of the mind wandering because we again discovered differences in ERP components when participants report being in on task or mind wandering states.

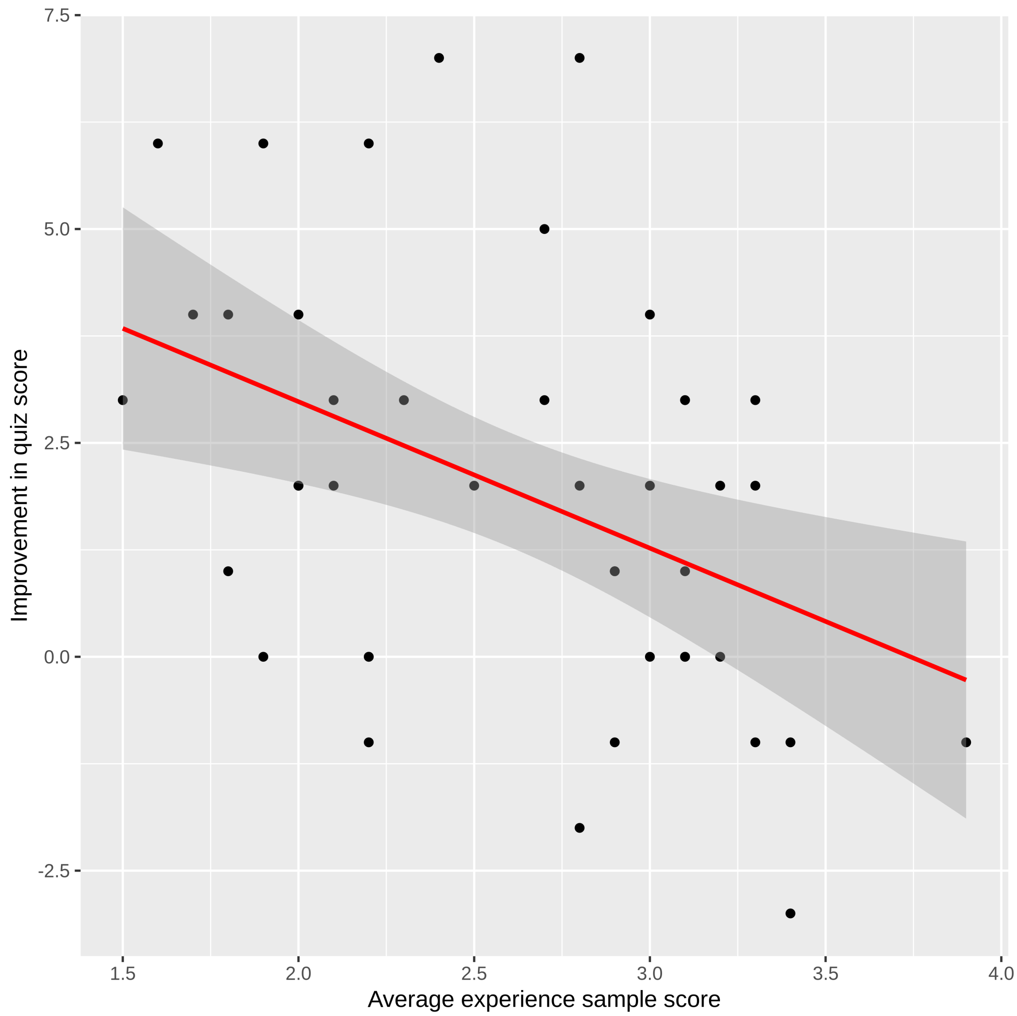


Figure 5.8: Best fit multivariate linear regression model of quiz score deltas

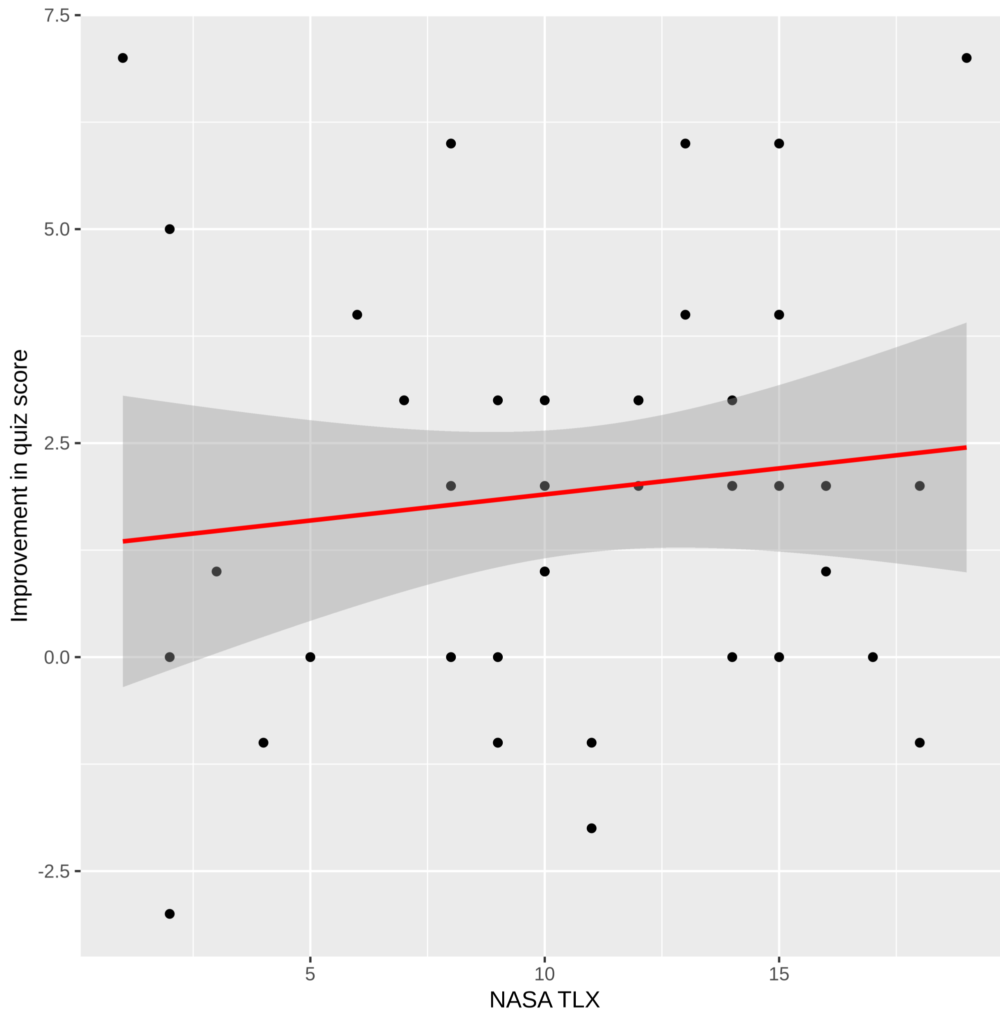


Figure 5.9: Linear regression of quiz score deltas and NASA TLX

P1 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	5.9574	0.3711	16.5	< 0.001***	
Amplitude	0.0157	0.3269	0.05	0.96	-0.0238

N1 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	5.8663	0.4912	11.94	< 0.001***	
Amplitude	-0.0601	0.2231	-0.27	0.79	-0.22

P2 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	5.9693	0.3696	16.15	< 0.001***	
Amplitude	-0.0675	0.2384	0.28	0.78	-0.219

P3 Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	5.9512	0.3730	15.96	< 0.001***	
Amplitude	-0.0114	0.2410	-0.05	0.96	-0.238

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.13: Summary of best fit linear regression models of quiz score deltas with EEG amplitudes

However, the components that differentiated the states were different from those discovered in the first experiment. Instead of an elevated P3 amplitude elicited by oddball stimuli during reported mind wandering, we discovered an association between N1 amplitude elicited by oddball stimuli during the on-task state. In addition, we discovered significant differences in P2 amplitude to standard stimuli between the two conditions. We are led to conclude that the different experience sampling methods employed during the two experiments resulted in different, yet useful ERP correlates. This potentially reflects two different EEG measures of mind wandering.

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	6.6806	1.4785	7.22	< 0.001***	
Exp. sample	-1.2480	0.5955	-2.10	0.042*	
Tech. rel. MW	-0.0687	0.0654	-1.05	0.300	
Tech. unr. MW	-0.0647	0.044	-1.47	0.149	
Whole Model	-	-	4.85	0.0057**	0.212

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.14: Summary of best fit multivariate linear regression model of quiz score deltas

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	3.256	0.618	5.27	< 0.001***	
Tech. rel. MW	–0.153	0.060	–2.53	0.015*	0.111

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.15: Summary of linear regression of differences in quiz scores (pre/post) and ex post reported technology related mind wandering

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	3.416	0.692	4.93	< 0.001***	
Tech. unr. MW	–0.105	0.043	–2.43	0.019*	0.102

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.16: Summary of linear regression of differences in quiz scores (pre/post) and reported technology unrelated mind wandering

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	6.407	1.499	4.28	< 0.001***	
Exp. sample	–1.712	0.562	–3.05	0.003**	0.161

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.17: Summary of linear regression of differences in quiz scores (pre/post) and average experience samples scores

Coefficients	Estimate	Std. Error	t or F score	p-value	Adj. R-sq.
(Intercept)	1.291	0.914	1.41	0.17	
NASA TLX	0.061	0.077	0.79	0.43	-0.009

† Significant at $\alpha = 0.1$; * Significant at $\alpha = 0.05$; ** Significant at $\alpha = 0.01$

Table 5.18: Summary of linear regression of differences in quiz scores (pre/post) and NASA TLX

In this experiment, elevated N1 and P2 responses when mind wandering suggest that participants oriented attention to the tones when on-task, but not when mind wandering. One explanation for this is that participants' attention was instead directed inwardly as their minds wandered and they tuned out the auditory tones. Though further research is warranted, the discovery of differences in the P2 response to standard stimuli could provide a useful measure of mind wandering that can have many applications in information systems research. By employing employing auditory tones alone and not relying on the oddball paradigm, researchers could potentially collect enough data to compare ERP responses in much less time.

At first glance, our discovery of differences in N1 and P2 amplitude seem to replicate some effects reported by Braboszcz and Delorme (2011). However, they reported elevated P2 responses when participants reported being in states of mind wandering, which is the opposite of what we found. Additionally, we did not manage to replicate the suggestive oscillatory results reported in Braboszcz and Delorme (2011) or those observed in Experiment 1. The small number of participants in the first experiment suggested it could have been the results of a false positive, and later revealed to be circumstantial with a larger number of participants. However, the fact that we did not observe the effects reported by Braboszcz and Delorme (2011) provides evidence that differences in oscillatory activity may not be generalizable and may be specific to the meditation task employed.

One possible reason why we did not observe a P3 response to the auditory tones is the difference in the experience sampling methods employed in the two studies. In the first study, participants were tasked with reporting when they experienced mind wandering. One explanation as to the presence of the P3 response to oddball stimuli is that when participants fell into a mind wandering state, their attention shifted towards the auditory sounds and pushing the button. In the second experiment, participants had no such task, so did not identify the stimuli as task-relevant. This may explain why we also observed lower amplitudes generally when in the mind wandering state in the second study.

Though we were somewhat surprised that there was no statistical relationship between average ERP amplitude and reported ex post mind wandering, these findings lend support to one of the original motivations for undertaking this research. Given the consistent correlation between on task states and N1 amplitude, for instance, we would expect that any similar measure to the mind wandering state would also share this relationship. However, we did not observe any such relationships. We did observe a relationship between aggregated state-level mind wandering and the ex post measures. We also observed a negative relationship between task load. Post hoc analysis also revealed suggestive ($p < 0.1$) relationship between EEG beta activity and task load.

The effects of gender and age on experience sample means was an interesting and surprising finding. Results indicated that participants identifying as female were significantly more likely to report mind wandering than males and that older participants reported less mind wandering. It is possible that these findings were the result of the video stimulus itself. Though women outnumbered men in this experiment (30 females, 14 males observed) the results are not necessarily effected by differences in the sample size. The video was very long (75 minutes) and featured a woman presenter. It is possible that these findings reflect gender bias toward the presenter, which has been reported in the literature (Centra & Gaubatz, 2000), or perhaps some other effect of gender. Given that the topic was venture capital, older participants may have been more apt to find the topic interesting due to more experience in the workplace. Though it is not definitive, these phenomena warrant further investigation.

Finally, it is noteworthy that though the the various mind wandering reports were correlated with rote learning, the task load survey measure was not. At face value, this challenges the Cognitive Theory of Multimedia learning, which posits that task load generated by extraneous factors inhibits learning. However, this lesson was not media-rich and participants were in a controlled environment with few extraneous factors. Future research could explore the relationship between mind wandering and cognitive load to potentially discover how these factors interact to inhibit or facilitate learning.

5.3.1 Limitations

We did not observe a relationship between ex post mind wandering and ERP amplitudes, but did observe relationships between experience sample means and ex post measures. This suggests that the observed ERP paradigm is currently limited, and must be refined to measure differences at the group level. At face value, this would not seem surprising, as differences between individuals are often more significant than differences between conditions (Luck, 2014). However, the linear mixed effects analysis employed considers variance between subjects. We are led to conclude that there is a gap between the neurophysiological measures employed and the survey measures employed in mind wandering research. We thus have evidence that neurophysiological measures lend useful insight, but are not merely the reduction of psychological measures. This is a considerable limitation to the potential of the use of ERP for measuring mind wandering in information systems settings. Information systems experiments should take care to consider fundamental differences between questionnaire and neurophysiological measures and use multiple measures in future studies. We discuss the consequences of this finding in Chapter 6.

A second challenge with these results is that the timing of the components was earlier than expected. Our technical tests suggested that the audio trigger delay caused an error that was approximately 52 ms later than it should have been. However, after correcting for the trigger delay, the potentials observed were revealed to be approximately 100 ms earlier than the first study. It is possible that this was caused by differences in the Windows operating environment's background applications between when the EEG recordings were made and when the audio timing test was conducted. However, it is also possible that this was caused by differences in the paradigms employed in the experiments. These concerns could be addressed by conducting another study that replicates the results of Experiment 2.

Another limitation is that we did not manage to replicate the electroencephalography (EEG) decoding results due to limitations in our experiment design. Given our priority of controlling for gaps between experience samples rather than number of samples, each participant yielded no more than 10 time samples that could be used for decoding. As such, we are limited in what we can conclude about the usefulness of this technique. Future work may overcome this problem by incorporating transfer learning, which may be able to overcome the challenges by transferring classification information between subject tasks (Pan & Yang, 2010). Future work could consider experiment designs which are appropriate for replicating the decoding results and perhaps apply the findings to real time analysis.

Finally, though we distinguished mind wandering which may be technology-related or technology unrelated, we did not distinguish mind wandering which may be task-related from that which is task unrelated. It is possible that some mind wandering may not inhibit learning if it is task related. Future work may explore different dimensions of the mind wandering construct and its affect on learning outcomes.

5.4 Summary

This chapter described a second experiment to measure the neurophysiological indicators of mind wandering, but also their relationship to learning efficacy. This experiment sought to overcome many of the limitations of Experiment 1 by employing a larger sample size, a different experience sampling method, and a recording correction. It employed ex post scales and quiz measures to assess the impact of mind wandering on learning. Observed results replicated some key findings of Experiment 1 but did not not replicate the P3 effect. Furthermore, we did not identify a relationship between the ERP or oscillatory measures and the ex post measures. However, we identified a clear relationship between mind wandering and learning. These findings suggest that there is a disconnect between mind wandering measured and real time and perceived mind wandering after the fact. In the final chapter, we will revisit the hypotheses described in Chapter 3 and draw conclusions about future research and applications.

Chapter 6

Conclusions

At the outset of this dissertation we discussed how Massive Open Online Courses (MOOCs) had, in many ways, failed to live up to their potential, and that one of the possible reasons for this is that they are not very good at keeping users' attention. We had three motivations for taking a neurophysiological approach to this question. The first is that we could learn more about attention-related information systems constructs by exploring neurophysiology. The second is that it could offer insight into the experience of mind wandering; unlike post-hoc questionnaires, we could potentially observe changes in real time. The third is that this research could open doors to the development of new types of technologies. We then formulated the following two underlying research questions:

RQ1 Can we use neurophysiological measures to detect differences in mind wandering during online lectures?

RQ2 Does mind wandering affect how well we learn from online lectures?

Upon revisiting these question, we are quickly led to a conclusion. Neurophysiological measures certainly can be used to measure attention; the literature had already established this. We sought to apply attention-related phenomena to measure mind wandering, a construct that has been previously observed in information systems research. We found that the the components of the P1-N1-P2 complex, especially N1-P2 and P3 amplitude, are significant correlates of on task or mind wandering states, but the markers uncovered were not reliable. We are thus left with two potential measures of mind wandering which could be further validated in future work. Our exploration also revealed that there is a gap between observed neurophysiological phenomena and questionnaire measures, especially in the way they measure group-level effects. We will discuss the implications of this research by revisiting the hypotheses outlined in

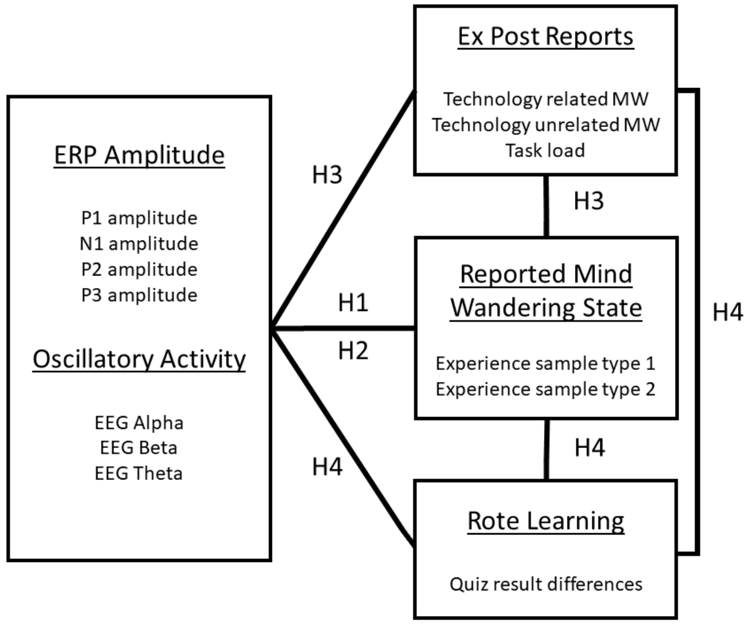


Figure 6.1: The proposed research model originally outlined in Chapter 3

Chapter 3. These hypotheses are organized to reflect the two questions described above.

6.1 Hypotheses Revisited

6.1.1 ERP Markers of Mind Wandering

This dissertation describes two experiments that were conducted in which EEG measures were correlated with reported mind wandering in an online lecture or MOOC setting. In both studies, we observed significant differences between event related potential amplitudes among the mind wandering and on task states. However, we observed two distinctive patterns which were unique to each study. In the first study, we significant differences in P1, P2 and P3 component amplitudes elicited by oddball stimuli versus standard stimuli during reported mind wandering states, but not when on-task. In the second study we observed significant differences in N1 and P2 component amplitude elicited by oddball stimuli versus standard stimuli during reported on task states, but not during reported mind wandering. Furthermore, in Experiment 2 we observed differences in P2 amplitudes elicited by standard stimuli during the mind wandering and on task states. We did not observe significant differences

Hypothesis	Finding	Description
H1a	supported	Mean P1 amplitude will distinguish on task and mind wandering states.
H1b	supported	Mean N1 amplitude will distinguish on task and mind wandering states.
H1c	supported	Mean P2 amplitude will distinguish on task and mind wandering states.
H1d	supported	Mean P3 amplitude will distinguish on task and mind wandering states.
H1e	not supported	Theta power will be positively correlated with the reported mind wandering state.
H1f	not supported	Alpha power will be positively correlated with the reported mind wandering state.
H1f	not supported	Beta power will be negatively correlated with the reported mind wandering state.
H2	not supported	There will be statistically significant correlations between neurophysiological measures and both types of experience samples.

Table 6.1: Summary of hypotheses and findings related to neurophysiological correlates of the mind wandering state during e-learning

in oscillatory activity between mind wandering and on-task states in either study. This suggests that differences in ERP amplitudes may be reliable measures of mind wandering during MOOCs, though further work is required to further elaborate these findings. However, we also have evidence that oscillatory patterns are not reliable measures.

It is likely that the differences between the ERP results observed in the studies reflect differences in the paradigms employed. In the first study, participants were asked to press a button when they experienced mind wandering. It is possible that the oddball tones triggered a P3b response because the participants became externally focused on the task of making the mind wandering report, and the tones became task-relevant. In the second study, participants had no such task and were prompted randomly. It is possible that participants instead focused their attention inward during the mind wandering state and tuned out external stimuli, such as the auditory tones.

We are thus led to question the scope and application of the EEG mind wandering measures that we chose. Though we observed differences in the P1-N1-P2 response in both studies, there were differences between the studies which resulted in differences in relevant neural markers. It is therefore likely that the amplitudes and timings of the ERPs will not generalize to other information technology tasks. However, it is possible that the general process of observing differences in attention-related potentials may generalize and could be the ultimate legacy of this work. Future research employing ERP for attention research in information systems should thus take care to similarly establish the measure being observed and replicate the result before drawing conclusions about technology phenomena.

These findings also offer evidence for the reliability of both types of experience sampling methods. Given that P1, N1, and P2 amplitudes have been associated with alerting and orienting attention networks (Gonçalves, Rêgo, et al., 2018), we have strong evidence that both the button press method described Braboszcz and Delorme (2011) and the Wammes and Smilek (2017) prompt method reliably measure shifts of attention towards and away from stimuli. We now know that the reported mind wandering state during an online lecture reflects attention processes.

Hypothesis	Finding	Description
H5	supported	An algorithm can be constructed which measures differences in reported mind wandering states with at least 70% accuracy.

Table 6.2: Summary of H5, a hypotheses related to the real-time measurement of the mind wandering state

Related to the contributions described above, we also succeeded at decoding real-time mind wandering brain signals, albeit using a different method than expected. At the outset, we expected that time-series pre-preparation would produce the strongest classification results. Motivated by work on the P300 speller, an attention-related classifier, we envisioned that time series analysis would produce strong results (Farwell & Donchin, 1988). However, we were unable to develop a time series classifier that performed better than random chance. Instead, frequency domain classifiers that classified better than random chance were produced. We demonstrated a classifier that used common spatial patterns and the SVM classifier to analyze data at the theta, alpha and beta frequencies (4-30 Hz). The selected classifier attained individual classification results ranging from 50% accuracy to 87.5% accuracy (mean 69.8% accuracy), depending on the participant. It is surprising that this classifier performed well while the time domain classifiers did not. Given that we discovered statistically significant differences in the P1, N1, P2 and P3 ERPs, but not in oscillatory activity, we would expect time series classifiers to perform best.

We have evidence that the classifier identifies oscillatory patterns that are lost when they are aggregated. Differences in oscillatory activity were often significant at the individual level but did not generalize to the group. The CSP processing algorithm maximizes the difference in frequencies detected among all epochs in each class. When used with SVM, the dimensions are reduced before the kernel is applied. It could be that the classifier maximizes differences that are present in many, but not all epochs, and classifies them accordingly. This would explain why we see such high variance between the participants. Regardless, these results indicate that it is possible to create a decoding technique that can reliably interpret differences in mind wandering in real-time. It is clear that differences in brain patterns are reliable indicators of the mind wandering state, and that we can build a classifier which can

Hypothesis	Finding	Description
H3a	not supported	Component amplitudes (P1, N1, P2, P3) will be positively correlated with reported technology unrelated mind wandering.
H3b	not supported	Component amplitudes (P1, N1, P2, P3) will be positively correlated with reported technology related mind wandering.
H3c	not supported	Component amplitudes (P1, N1, P2, P3) will be negatively correlated with reported task load.
H3d	not supported	Reported mind wandering experiences will be positively correlated with reported technology unrelated mind wandering.
H3e	supported	Reported mind wandering experiences will be positively correlated with reported technology related mind wandering.
H3f	supported	Reported mind wandering experiences will be negatively correlated with reported task load.

Table 6.3: Summary of hypotheses and findings related to EEG and ex post mind wandering

reliably detect real time differences in mind wandering.

6.1.2 Mind Wandering Consists of More than Attention

Though we observed significant differences in ERPs between the mind wandering and on-task states, these results did not generalize to the ex post measures. We did not observe any significant relationship between questionnaires and component amplitudes. However, we did observe a significant relationship between the average of experience samples and the ex post questionnaires.

We are therefore led to conclude that there is a gap between observe neurophysiological indicators and the mind wandering construct being reported. At the outset of this research, we associated mind wandering with attention processes and discovered positive correlations between the neurophysiological signals which are associated

Hypothesis	Finding	Description
H4a	not supported	Quiz deltas will be negatively correlated with component amplitudes.
H4b	supported	Quiz deltas will be negatively correlated with reported technology unrelated mind wandering.
H4c	supported	Quiz deltas will be negatively correlated with reported technology related mind wandering.
H4d	not supported	Quiz deltas will be negatively correlated with reported task load.
H4e	supported	Quiz deltas will be negatively correlated with reported mind wandering states.

Table 6.4: Summary of hypotheses and findings related to group-level ex post mind wandering and rote learning measures

with attention and mind wandering. However, mind wandering is also associated with other mental processes such as task-unrelated thoughts and creativity (Sullivan et al., 2015). The neurophysiological measures employed by this study are not correlates of these elements of mind wandering, which may confound the transitivity between observations and the psychological measures employed. Future work may wish to incorporate different neurophysiological measures to better capture those dimensions of mind wandering.

6.1.3 The Small Effect of Mind Wandering on Learning

In both studies participants performed significantly better on the multiple choice quiz following the lecture video versus the quiz conducted before. However, did not observe a relationship between ERP component amplitudes and quiz delta, but did observe statistically significant relationships between quiz delta, mean experience samples and the ex post scales. In the best fit linear regression model of quiz delta, all mind wandering measures had negative coefficients. We are thus led to the conclusion that mind wandering has a negative impact on online learning efficacy. However, this relationship is likely not the largest factor in MOOC efficacy. With an R-squared value just over 0.2, it is likely that there are larger factors in how well participants learn from videos. We conclude that MOOCs should consider designs that reduce mind wandering, but not at the expense of other facts which influence e-learning success.

6.2 Contributions and Future Work

6.2.1 ERP Attention Measures for Information Systems Research

Perhaps the greatest contribution of this dissertation is the demonstration of a passive, real-time measure of mind wandering during technology use. By observing the differences in P1, N1, P2 and P3 amplitudes, we can reliably measure the on-task and mind wandering states. This discovery has a number of potential applications to IS research which should be highlighted.

The first is the application of this passive mind wandering measure to other IS artifacts. Though mind wandering is particularly relevant to MOOCs, it is also applicable to other domains such as serious games, business analytics, design science, and e-commerce. By measuring differences in the N1, P2 and P3 amplitudes elicited by auditory oddball stimuli, we can gain new insights into whether mind wandering impacts the outcomes of such artifacts. Additionally, EEG decoding can be employed to learn about changes in the interaction with IS artifacts over time.

The second is in its applicability to other attention-related IS constructs. Though we demonstrated neurophysiological correlates of mind wandering, the P1-N1-P2 complex is fundamentally the product of attention processes in the brain. Other constructs such as acceptance, adoption, communication, user satisfaction, flow or decision making all reflect dimensions of attention. Though IS constructs do not necessarily reduce to atomic or measurable neurophysiological measures, as demonstrated in this dissertation, the event related potentials described in this dissertation could be used to discover more about the relationship between attention and other IS constructs.

Finally, the EEG decoding technique demonstrated in this paper can also be employed to offer new insight into how brain states change over time and can be used to develop new types of neuro-adaptive information technologies. To the best of our knowledge, no IS research has employed EEG decoding techniques to reveal real-time changes in brain states. Future work could employ the Common Spatial Patterns technique to construct classifiers that offer insight into this dimension of IS phenomena. It can also apply the technique to generate e-learning brain-computer intferface (BCI)s that change with brain states, potentially offering a superior learning experience. Future work may improve on the results described in this dissertation by applying transfer learning to these problems and ultimately overcome the need for a large number of mind wandering report samples (Gayraud et al., 2017). One potential ways to improve the paradigm could include the experience sampling rate or to collect data from a longer video.

6.2.2 How to Create Better Teaching Techniques

A second contribution of this research is that it extends the discussion about the role that mind wandering plays in the online classroom. We ultimately found that mind wandering has an impact on rote learning. Educators with the goal to transfer factual knowledge can take heed of these findings and discover ways to effectively design lectures, either online or in-person, in a way that limits mind wandering. Some researchers have suggested that asking questions may help manage mind wandering (Szpunar et al., 2013) while others have noted that active learning approaches offer promise for improving MOOC media (Ubell, 2017).

We should qualify that though we observed a relationship between mind wandering and rote learning, there are likely other factors in online lecture efficacy and likely other factors in MOOCs specifically. For instance, a recent study in MOOC completion found that 79% of the variance in completion is determined by the quality of content, degree of interaction with the instructor and perceived effectiveness of a MOOC (Hone & El Said, 2016). Future studies on the impact of mind wandering in MOOCs should also consider these factors in addition to mind wandering. They should also consider the interaction effects between mind wandering and these factors. By doing so, we can more precisely identify the role that mind wandering or similar attentional processes play in MOOC effectiveness. Flow, active learning and multimedia are all related subjects of inquiry which might be further investigated using the techniques described in this dissertation.

In addition, this dissertation explored the impact of mind wandering on rote learning in the context of a single lecture. Rote learning is only one possible learning outcome and is often not the goal of quality higher education. As mentioned near the beginning of this dissertation, what Mayer calls *meaningful learning* is a more desirable goal and often takes place over an extended period of time (Mayer, 2009). Future work could consider the effects of persistent mind wandering over a period of time, especially in the acquisition of meaningful learning outcomes. Such research would better identify the role that mind wandering plays in MOOC effectiveness, and could potentially offer insight into the development of effective, low cost learning technologies.

Finally, the findings of this research are limited by the fact that the IT artifacts explored were not manipulated. Though there is evidence that mind wandering has an impact on MOOC efficacy, we did not identify a method for improving MOOC multimedia. Future work can be conducted on teaching techniques which limit mind wandering and can identify effective multimedia for improving learning outcomes. Researchers could consider investigating how video length or the presence of in-lecture quizzes impact mind wandering and learning outcomes.

6.2.3 On the Reduction of Constructs into Neurophysiology

A third philosophical finding of this dissertation not directly reflected by the hypotheses is the discovery of a gap between neurophysiological measures and ex post questionnaire measures of mind wandering. In this case, the P1-N1-P2 complex components consistently reflect differences in the mind wandering state, but are are not reflective of ex post mind wandering reports. We can interpret these results to suggest that there is a difference between participants' perception of task unrelated thoughts and their experience of them. However, there are consequences of these results.

In a paper presented at the 2018 NeuroIS retreat, researchers suggested that this gap between neurophysiological measures and questionnaire instruments offers evidence for the philosophical position of mind body dualism (Buettner, Bachus, Konzmann, & Prohaska, 2019). This theory, most famously advanced by Réné Descartes, suggests that mind and body are fundamentally different from one another, insofar as they consist of different metaphysical substances. At face value, the results described in this dissertation appear to lend support for this theory. Though physiological measures were correlated with reports at the individual level, they were not correlated with instruments that had previously been found to be significant predictors of mind wandering.

There are other potential explanations for these results however. The phenomenon being observed is mind wandering, a somewhat poorly defined construct used to describe a set of experiences in ways that are easy for humans to understand. The neurophysiological correlates being observed are reflections of well-defined attention processes, but are merely correlates of the questionnaire constructs explored. When comparing constructs we pass the buck of causality to human reasoning. Even path models, often used to model causal relationships in the social sciences, are only as good as the causal models that humans construct, and are not derived from data alone (Pearl, 2009). By further refining the paradigm, we might yet overcome these limitations and eventually get better results at the group level.

We argue that this gives evidence that the ex post questionnaires are simply not reducable to the neurophysiological indicators explored in these studies. EEG (like other neuroimaging technologies) measures aggregates of signals produced by millions of neurons. In the case of EEG, we measure electrical signals that must pass through a conductive skull. It should not be surprising that the relationships observed are not correlated with ex post questionnaire measures, which measure perceptions of experiences. Future work in NeuroIS should consider the domain differences in theories and research methods employed and that there is no a priori reason why existing IS constructs should reduce to observed neurophysiological phenomena observed using neuroimaging technologies. By further improving the paradigm employed through additional studies, we may yet overcome these challenges. We believe that this dissertation may offer a small part in a larger discussion about the role that neuroscience has to play in extending the field of information systems moving forward, and how the methods employed may yet be improved to overcome perceived limitations.

6.3 Summary

This dissertation had two overarching goals. The first was to identify a neurophysiological measure that can detect differences in mind wandering during online lectures.

The second was to identify whether mind wandering affects how we learn from online lectures. We successfully identified how brain responses to oddball stimuli are correlates of on task and mind wandering states in an online learning setting. Though we did not identify significant relationships between physiological responses and rote learning outcomes, we nonetheless identified negative relationship between mind wandering and rote learning. This relationship explained just over 20% of the variance in learning however. We were thus led to conclude that mind wandering has a small but significant impact on learning. MOOC providers should therefore consider designs that limit mind wandering, but not at the expense of other design factors which may influence learning outcomes.

References

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, *24*, 665–694.

Akram, F., Han, S. M., & Kim, T.-S. (2015). An efficient word typing P300-BCI system using a modified t9 interface and random forest classifier. *Computers in Biology and Bedicine*, *56*, 30–36.

Anaconda. (2019). *Anaconda distribution: The world's most popular Python/R data science platform.* Retrieved from `https://www.anaconda.com/distribution/` (Online; accessed March 26 2019)

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Baker, D. (2013). *Arduino sound to TTL trigger for EEG.* Retrieved from `https://bakerdh.wordpress.com/` (Online; accessed January 3 2019)

Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lerner, N., & Higgins, J. S. (2017). Detecting and quantifying mind wandering during simulated driving. *Frontiers in Human Neuroscience*, *11*, 406.

Bashashati, A., Fatourechi, M., Ward, R. K., & Birch, G. E. (2007). A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, *4*(2), R32.

Beane, M., & Marrocco, R. (2004). Norepinephrine and acetylcholine mediation of the components of reflexive attention: implications for attention deficit disorders. *Progress in Neurobiology*, *74*(3), 167–181.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, M. N., ... Olmstead, R. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction*, *17*(2), 151–170.

Boutin, K.-D., Léger, P.-M., Davis, C. J., Hevner, A. R., & Labonté-LeMoyne, É. (2019). Attentional characteristics of anomaly detection in conceptual modeling.

In *Information systems and neuroscience* (pp. 57–63). Springer.

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: neural markers of low alertness during mind wandering. *Neuroimage*, *54*(4), 3040–3047.

Brandmeyer, T., & Delorme, A. (2018). Reduced mind wandering in experienced meditators and associated EEG correlates. *Experimental brain research*, *236*(9), 2519–2528.

Brocke, J. V., Riedl, R., & Léger, P.-M. (2013). Application strategies for neuroscience in information systems design science research. *Journal of Computer Information Systems*, *53*(3), 1–13.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 53–61.

Buettner, R., Bachus, L., Konzmann, L., & Prohaska, S. (2019). Asking both the users heart and its owner: Empirical evidence for substance dualism. In *Information systems and neuroscience* (pp. 251–257). Springer.

Calic, G., El Shamy, N., Hassanein, K., & Watter, S. (2018). The impact of age and cognitive style on e-commerce decisions: The role of cognitive bias susceptibility. In *Information systems and neuroscience.* Springer.

Callejas, A., Lupianez, J., Funes, M. J., & Tudela, P. (2005). Modulations among the alerting, orienting and executive control networks. *Experimental Brain Research*, *167*(1), 27–37.

Cecotti, H., & Graser, A. (2011). Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(3), 433–445.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education*, *71*(1), 17–33.

Christoff, K., Gordon, A., Smith, R., & Vancouver, B. (2011). The role of spontaneous thought in human cognition. *Neuroscience of Decision Making*, 259–284.

Clark, R. C., Mayer, R. E., & Thalheimer, W. (2003). E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning. *Performance Improvement*, *42*(5), 41–43.

Cole, M. J., Hendahewa, C., Belkin, N. J., & Shah, C. (2015). User activity patterns during information search. *ACM Transactions on Information Systems (TOIS)*, *33*(1), 1.

Congedo, M., Goyat, M., Tarrin, N., Ionescu, G., Varnet, L., Rivet, B., . . . Jutten, C. (2011). " brain invaders": a prototype of an open-source P300-based video game working with the openvibe platform. In *5th international brain-computer interface conference 2011 (BCI 2011)* (pp. 280–283).

Conrad, C., Bliemel, M., & Ali-Hassan, H. (2019). The role of flow in learning distributed computing and MapReduce concepts using hands-on analogy. *Journal of Information Systems Education*, *30*(1), 57.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., ... Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761–773.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Review of Neuroscience*, *3*(3), 201.

Courtemanche, F., Fredette, M., Senecal, S., Leger, P.-M., Dufresne, A., Georges, V., & Labonte-lemoyne, E. (2018, February 8). *Method of and system for processing signals sensed from a user.* Google Patents. (US Patent App. 15/552,788)

Csikszentmihalyi, M. (2014). Toward a psychology of optimal experience. In *Flow and the foundations of positive psychology* (pp. 209–226). Springer.

Cyr, D., Head, M., Larios, H., & Pan, B. (2009). Exploring human images in website design: a multi-method approach. *MIS Quarterly*, 539–566.

Dabbish, L., & Kraut, R. (2008). Research noteawareness displays and social motivation for coordinating communication. *Information Systems Research*, *19*(2), 221–238.

Davidson, D. J. (2009). Functional mixed-effect models for electrophysiological responses. *Neurophysiology*, *41*(1), 71–79.

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

Delorme, A., Miyakoshi, M., Jung, T.-P., & Makeig, S. (2015). Grand average erp-image plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions. *Journal of Neuroscience Methods*, *250*, 3–6.

De Manzano, Ö., Theorell, T., Harmat, L., & Ullén, F. (2010). The psychophysiology of flow during piano playing. *Emotion*, *10*(3), 301.

Descartes, R. (1984). *The philosophical writings of descartes: Volume 3, the correspondence* (Vol. 3; D. M. John Cottingham Robert Stoothoff & A. Kenny, Eds.). Cambridge University Press.

Dijksterhuis, A., & Meurs, T. (2006). Where creativity resides: The generative power of unconscious thought. *Consciousness and Cognition*, *15*(1), 135–146.

Dimoka, A., & Davis, F. D. (2008). Where does tam reside in the brain? the neural mechanisms underlying technology adoption. *ICIS 2008 Proceedings*, 169.

Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., ... Weber, B. (2012). On the use of neurophysiological tools in is research: Developing a research agenda for neuroIS. *MIS Quarterly*, 679–702.

Dimoka, A., Pavlou, P. A., & Davis, F. D. (2011). Research commentaryneuroIS: The potential of cognitive neuroscience for information systems research. *Information Systems Research*, *22*(4), 687–702.

Djamasbi, S., Siegel, M., & Tullis, T. (2012). Designing noticeable bricklets by tracking users' eye movements. In *2012 45th hawaii international conference on system science* (pp. 525–532).

Donchin, E., Spencer, K. M., & Wijesinghe, R. (2000). The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, *8*(2), 174–179.

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105.

Drescher, L. H., Van den Bussche, E., & Desender, K. (2018). Absence without leave or leave without absence: Examining the interrelations among mind wandering, metacognition and cognitive control. *PloS one*, *13*(2), e0191639.

Dunchin, E., & Heffley, E. F. (1978). Multivariate analysis of event-related potential data: A tutorial review. In *Multidisciplinary perspectives in event-related brain potential research* (p. 555-572). U.S. Government Printing Office.

Edlinger, G., Holzner, C., & Guger, C. (2011). A hybrid brain-computer interface for smart home control. In *International conference on human-computer interaction* (pp. 417–426). Springer.

El Shamy, N., & Hassanein, K. (2018). The impact of age and cognitive style on e-commerce decisions: The role of cognitive bias susceptibility. In *Information systems and neuroscience* (pp. 73–83). Springer.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23.

Fan, J., Gu, X., Guise, K. G., Liu, X., Fossella, J., Wang, H., & Posner, M. I. (2009). Testing the behavioral interaction and integration of attentional networks. *Brain and Cognition*, *70*(2), 209–220.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *Neuroimage*, *26*(2), 471–479.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, *70*(6), 510–523.

Fazel-Rezai, R., Allison, B. Z., Guger, C., Sellers, E. W., Kleih, S. C., & Kübler, A. (2012). P300 brain computer interface: current challenges and emerging trends. *Frontiers in Neuroengineering*, *5*.

Firmin, R., Schiorring, E., Whitmer, J., Willett, T., Collins, E. D., & Sujitparapitaya, S. (2014). Case study: Using MOOCs for conventional college coursework. *Distance Education*, *35*(2), 178–201.

Fischer, T., Davis, F. D., & Riedl, R. (2019). NeuroIS: A survey on the status of the field. In *Information systems and neuroscience* (pp. 1–10). Springer.

Flach, P. (2012). *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410–8415.

Fu, E. (2017). *Introduction to venture capital.* Retrieved from `https://www.youtube.com/watch?v=L8NOXl6Taegt=3860s` (Online; accessed January 3 2019)

Furdea, A., Ruf, C., Halder, S., De Massari, D., Bogdan, M., Rosenstiel, W., ... Birbaumer, N. (2012). A new (semantic) reflexive brain–computer interface: In search for a suitable classifier. *Journal of Neuroscience Methods*, *203*(1), 233–240.

Gayraud, N. T., Rakotomamonjy, A., & Clerc, M. (2017). Optimal transport applied to transfer learning for P300 detection. In *7th graz brain-computer interface conference 2017.*

Geva, R., Zivan, M., Warsha, A., & Olchik, D. (2013). Alerting, orienting or executive attention networks: differential patters of pupil dilations. *Frontiers in Behavioral Neuroscience*, *7*, 145.

Gonçalves, Ó. F., Carvalho, S., Mendes, A. J., Leite, J., & Boggio, P. S. (2018). Neuromodulating attention and mind-wandering processes with a single session real time EEG. *Applied Psychophysiology and Biofeedback*, *43*(2), 143–151.

Gonçalves, Ó. F., Rêgo, G., Conde, T., Leite, J., Carvalho, S., Lapenta, O. M., & Boggio, P. S. (2018). Mind wandering and task-focused attention: Erp correlates. *Scientific Reports*, *8*(1), 7608.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Mamalainen, M. (2013). Meg and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460.

Gregori, E. B., Zhang, J., Galván-Fernández, C., & de Asís Fernández-Navarro, F. (2018). Learner support in MOOCs: Identifying variables linked to completion. *Computers & Education*, *122*, 153–168.

Grimson, E. (2017). *Introduction to machine learning.* Retrieved from `https://www.youtube.com/watch?v=h0e2HAPTGF4t=428s` (Online; accessed August 8 2018)

Gu, Z., Yu, Z., Shen, Z., & Li, Y. (2013). An online semi-supervised brain–computer interface. *IEEE Transactions on Biomedical Engineering*, *60*(9), 2614–2623.

Guger, C., Daban, S., Sellers, E., Holzner, C., Krausz, G., Carabalona, R., ... Edlinger, G. (2009). How many people are able to control a P300-based brain–computer interface (BCI)? *Neuroscience Letters*, *462*(1), 94–98.

Gwizdka, J. (2019). Exploring eye-tracking data for detection of mind-wandering on web tasks. In *Information systems and neuroscience* (pp. 47–55). Springer.

Haas, L. F. (2003). Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(1), 9–9.

Harding, R. J. (2019). Open notebook science can maximize impact for rare disease projects. *PLoS biology*, *17*(1).

Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *Journal of Neuroscience*, *33*(7), 2927–2933.

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139–183.

Haslam, K., Doucette, H., Hachey, S., MacCallum, T., Zwicker, D., Smith-Brilliant, M., & Gilbert, R. (2019). Youtube videos as health decision aids for the public: An integrative review. *Canadian Journal of Dental Hygiene*, *53*(1).

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456.

Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences*, *95*(3), 781–787.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180.

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *353*(1373), 1257–1270.

Hoffman, D. L., & Novak, T. P. (2009). Flow online: lessons learned and future prospects. *Journal of Interactive Marketing*, *23*(1), 23–34.

Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting mooc retention: A survey study. *Computers & Education*, *98*, 157–168.

Hong, W., Hess, T. J., & Hardin, A. (2013). When filling the wait makes it feel longer: A paradigm shift perspective for managing online delay. *MIS Quarterly*, *37*(2), 383–406.

Hong, W., Thong, J. Y., & Tam, K. Y. (2004). Does animation attract online users attention? the effects of flash on information search performance and perceptions. *Information Systems Research*, *15*(1), 60–86.

Horváth, J., & Burgyán, A. (2011). Distraction and the auditory attentional blink. *Attention, Perception, & Psychophysics*, *73*(3), 695–701.

Hsu, J. (2012). Professor leaving Stanford for online education startup. *NBC News*. (Accessed: 2018-06-28)

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, *9*(3), 90.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE transactions on Neural Networks*, *10*(3), 626–634.

Ishigami, Y., Eskes, G. A., Tyndall, A. V., Longman, R. S., Drogos, L. L., & Poulin, M. J. (2016). The attention network test-interaction (ANT-I): reliability and validity in healthy older adults. *Experimental Brain Research*, *234*(3), 815–827.

Ishigami, Y., & Klein, R. M. (2009). Are individual differences in absentmindedness correlated with individual differences in attention? *Journal of Individual Differences*, *30*(4), 220–237.

Ishigami, Y., & Klein, R. M. (2011). Repeated measurement of the components of attention of older adults using the two versions of the attention network test: stability, isolability, robustness, and reliability. *Frontiers in Aging Neuroscience*, *3*, 17.

Jackson, S. A., & Eklund, R. C. (2004). *The flow scales manual*. Fitness Information Technology.

James, W. (1890). *The principles of psychology* (Vol. 2). Henry Holt and Company.

Janiszewski, C. (1998). The influence of display characteristics on visual exploratory search behavior. *Journal of Consumer Research*, *25*(3), 290–301.

Jensen, O., Kaiser, J., & Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317–324.

Jones, S. A., Butler, B. C., Kintzel, F., Johnson, A., Klein, R. M., & Eskes, G. A. (2016). Measuring the performance of attention networks with the dalhousie computerized attention battery (DalCAB): Methodology and reliability in healthy adults. *Frontiers in Psychology*, *7*.

Jung, Y., & Lee, J. (2018). Learning engagement and persistence in massive open online courses (MOOCS). *Computers & Education*, *122*, 9–22.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Citeseer.

Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., & Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an erp (P300) brain–computer interface. *Biological Psychology*, *102*, 118–129.

Klein, R., Kingstone, A., & Pontefract, A. (1992). Orienting of visual attention. In *Eye movements and visual cognition* (pp. 46–65). Springer.

Klein, R. M. (2004). On the control of visual orienting. In M. I. Posner (Ed.), *Cognitive neuroscience of attention*. Guilford Press.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *IJCAI* (Vol. 14, pp. 1137–1145).

Koles, Z. J. (1991). The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG. *Electroencephalography and Clinical Neurophysiology*, *79*(6), 440–447.

Kolowich, S. (2013). MOOCs are largely reaching privileged learners, survey finds. *The Cronicle of Higher Education*. (Accessed: 2018-06-28)

Konnikova, M. (2014, November 7). Will MOOCs be flukes? *The New Yorker*. (Accessed: 2018-06-28)

Krusienski, D. J., Sellers, E. W., Cabestaing, F., Bayoudh, S., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2006). A comparison of classification techniques for the P300 speller. *Journal of Neural Engineering*, *3*(4), 299.

Lee, Y., Chen, A. N., & Ilie, V. (2012). Can online wait be managed? the effect of filler interfaces and presentation modes on perceived waiting time online. *MIS Quarterly*, 365–394.

Léger, P.-M., Davis, F. D., Cronan, T. P., & Perret, J. (2014). Neurophysiological correlates of cognitive absorption in an enactive training context. *Computers in Human Behavior*, *34*, 273–283.

Léger, P.-M., Davis, F. D., Perret, J., & Dunaway, M. (2010). Psychophysiological measures of cognitive absorption. *Proceedings of SIGHCI*.

Léger, P.-M., Sénecal, S., Courtemanche, F., de Guinea, A. O., Titah, R., Fredette, M., & Labonte-LeMoyne, É. (2014). Precision is in the eye of the beholder: Application of eye fixation-related potentials to information systems research. *Journal of the Association for Information Systems*, *15*(10), 651.

Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072.

Lin, F.-R., & Kao, C.-M. (2018). Mental effort detection using EEG data in e-learning contexts. *Computers & Education*, *122*, 63–79.

Lindquist, S. I., & McLean, J. P. (2011). Daydreaming and its correlates in an educational environment. *Learning and Individual Differences*, *21*(2), 158–167.

Liu, H.-C., Lai, M.-L., & Chuang, H.-H. (2011). Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers cognitive processes. *Computers in Human Behavior*, *27*(6), 2410–2417.

Locke, J. (1689). *An essay concerning human understanding* (1979th ed.; P. Nidditch, Ed.).

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, *15*(3), 031005.

Lu, J., Hallinger, P., & Showanasai, P. (2014). Simulation-based learning in management education: A longitudinal quasi-experimental evaluation of instructional effectiveness. *Journal of Management Development*, *33*(3), 218–244.

Luck, S. J. (2014). *An introduction to the event-related potential technique, 2nd edition.* MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any erp experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400.

Mayer, R. E. (2009). *Multimedia learning (2nd).* Cambridge University Press New York.

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction*, *29*, 171–173.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222.

McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (p. 51 - 56).

McVay, J. C., Kane, M. J., & Kwapil, T. R. (2009). Tracking the train of thought from the laboratory into everyday life: An experience-sampling study of mind wandering across controlled and ecological contexts. *Psychonomic Bulletin & Review*, *16*(5), 857–863.

Michalczyk, S., Jung, D., Nadj, M., Knierim, M. T., & Rissler, R. (2019). brownieR: The r-package for neuro information systems research. In *Information systems and neuroscience* (pp. 101–109). Springer.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Mitchell, C. J., & Le Pelley, M. E. (2010). *Attention and associative learning: From brain to behaviour.* Oxford University Press, USA.

Mole, C. (2017). Attention. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2017 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/fall2017/entries/attention/`.

Montemayor, C., & Haladjian, H. H. (2015). *Consciousness, attention, and conscious attention*. MIT Press.

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: a review. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 11.

Niebert, K., Marsch, S., & Treagust, D. F. (2012). Understanding needs embodiment: A theory-guided reanalysis of the role of metaphors and analogies in understanding science. *Science Education*, *96*(5), 849–877.

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. Psychology Press.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345–1359.

Pashler, H., Johnston, J. C., & Ruthruff, E. (2001). Attention and performance. *Annual review of psychology*, *52*(1), 629–651.

Pearce, J. M., Ainley, M., & Howard, S. (2005). The ebb and flow of online learning. *Computers in Human Behavior*, *21*(5), 745–771.

Pearl, J. (2009). *Causality*. Cambridge University Press.

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

Peirce, J. W. (2007). PsychoPy psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13.

Percival, D. B., & Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge University Press.

Perez, F., & Granger, B. E. (2015). Project Jupyter: Computational narratives as the engine of collaborative data science. *iPython Archive*, *11*(207), 108.

Perna, L., Ruby, A., Boruch, R., Wang, N., Scull, J., Evans, C., & Ahmad, S. (2013). The life cycle of a million mooc users. In *Mooc research initiative conference* (pp. 5–6).

Petersen, S., Fox, P., Miezin, F., & Raichle, M. (1988). Modulation of cortical visual responses by direction of spatial attention measured by pet. *Association for Research in Vision and Opthalmology, Abstracts*, *22*.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89.

Petley, L., Bardouille, T., Chiasson, D., Froese, P., Patterson, S., Newman, A., ... Beyea, S. (2018). Attentional dysfunction and recovery in concussion: effects on the P300m and contingent magnetic variation. *Brain Injury*, *32*(4), 464–473.

Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students patterns of engagement and course performance in a massive open online course. *Computers & Education*, *95*, 36–44.

Phillips-Wren, G. E., Iyer, L. S., Kulkarni, U. R., & Ariyachandra, T. (2015). Business analytics in the context of big data: A roadmap for research. *CAIS*, *37*, 23.

Polich, J. (2007). Updating P300: an integrative theory of p3a and p3b. *Clinical Neurophysiology*, *118*(10), 2128–2148.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, *40*(1), 187–195.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Posner, M. I. (1987). Structures and functions of selective attention. In T. Boll & B. K. Bryant (Eds.), *The master lecture series, vol. 7. clinical neuropsychology and brain function: Research, measurement, and practice* (p. 173-202). ERIC.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25–42.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, *58*, 1–23.

Potter, R. E., & Balthazard, P. (2004). The role of individual memory and attention processes during electronic brainstorming. *MIS Quarterly*, *28*, 621–643.

Quitadamo, L., Cavrini, F., Sbernini, L., Riillo, F., Bianchi, L., Seri, S., & Saggio, G. (2017). Support vector machines to detect physiological patterns for EEG and emg-based human–computer interaction: a review. *Journal of Neural Engineering*, *14*(1).

Rakotomamonjy, A., & Guigue, V. (2008). BCI competition iii: dataset ii-ensemble of svms for BCI P300 speller. *IEEE Transactions on Biomedical Engineering*, *55*(3), 1147–1154.

Randolph, A. B., Labonté-LeMoyne, É., Léger, P.-M., Courtemanche, F., Sénécal, S., & Fredette, M. (2015). Proposal for the use of a passive BCI to develop a neurophysiological inference model of is constructs. In *Information systems and neuroscience* (pp. 175–180). Springer.

Recker, J. (2012). *Scientific research in information systems: a beginner's guide.* Springer Science & Business Media.

Riedl, R., Davis, F. D., Banker, R., & Kenning, P. (2017). *Neuroscience in information systems research.* Springer.

Riedl, R., Fischer, T., & Léger, P.-M. (2017). A decade of neuroIS research: status quo, challenges, and future directions. In *Proceedings of the 38th international conference on information systems.*

Riedl, R., & Léger, P.-M. (2016). *Fundamentals of neuroIS.* Springer.

Riedl, R., Randolph, A. B., vom Brocke, J., Léger, P.-M., & Dimoka, A. (2010). The potential of neuroscience for human-computer interaction research. In *Proceedings of the 9th annual workshop on hci research in mis* (pp. 1–5).

Rivet, B., Souloumiac, A., Attina, V., & Gibert, G. (2009). xDAWN algorithm to enhance evoked potentials: application to brain–computer interface. *IEEE Transactions on Biomedical Engineering*, *56*(8), 2035–2043.

Roberts, B. M., Hsieh, L.-T., & Ranganath, C. (2013). Oscillatory activity during maintenance of spatial and temporal information in working memory. *Neuropsychologia*, *51*(2), 349–357.

Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple-choice testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1155.

Salvaris, M., & Sepulveda, F. (2009). Visual modifications on the P300 speller BCI paradigm. *Journal of Neural Engineering*, *6*(4), 046011.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229.

Schooler, J. W. (2004). Zoning out while reading: Evidence for dissociations between experience and metaconsciousness jonathan w. schooler, erik d. reichle, and david v. halpern. In *Thinking and seeing: Visual metacognition in adults and children* (pp. 203–226).

Schwab, K. (2017). *The fourth industrial revolution.* Crown Business.

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going awol in the brain: Mind wandering reduces cortical analysis of external events. *Journal of Cognitive Neuroscience*, *20*(3), 458–469.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, *13*(4), 657–690.

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*, 487–518.

Smith, D. T., & Schenk, T. (2012). The premotor theory of attention: time to move on? *Neuropsychologia*, *50*(6), 1104–1114.

Speier, W., Fried, I., & Pouratian, N. (2013). Improved P300 speller performance using electrocorticography, spectral features, and natural language processing. *Clinical Neurophysiology*, *124*(7), 1321–1328.

Sullivan, Y., Davis, F., & Koh, C. (2015). Exploring mind wandering in a technological setting. In *Proceedings of the 36th international conference on information systems, forth worth.*

Sutton, S., Braren, M., Zubin, J., & John, E. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296.

Szpunar, K., Moulton, S., & Schacter, D. (2013). Mind wandering and education: from the classroom to online learning. *Frontiers in Psychology*, *4*, 495. doi: 10.3389/fpsyg.2013.00495

Tallon-Baudry, C., Kreiter, A., & Bertrand, O. (1999). Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. *Visual Neuroscience*, *16*(3), 449–459.

Tam, K. Y., & Ho, S. Y. (2006). Understanding the impact of web personalization on user information processing and decision outcomes. *MIS Quarterly*, *30*, 865–890.

Tan, B. C., Yi, C., & Chan, H. C. (2008). Beyond annoyance: the latent benefits of distracting website features. *ICIS 2008 Proceedings*, 188.

Teder-Sälejärvi, W. A., Münte, T. F., Sperlich, F.-J., & Hillyard, S. A. (1999). Intra-modal and cross-modal spatial attention to auditory and visual stimuli. an event-related brain potential study. *Cognitive Brain Research*, *8*(3), 327–343.

Thompson, K. G., Biscoe, K. L., & Sato, T. R. (2005). Neuronal basis of covert spatial attention in the frontal eye field. *Journal of Neuroscience*, *25*(41), 9479–9487.

Thulasidas, M., Guan, C., & Wu, J. (2006). Robust classification of EEG signal for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*(1), 24–29.

Toelch, U., & Ostwald, D. (2018). Digital open scienceteaching digital tools for reproducible and transparent research. *PLoS biology*, *16*(7).

Treagust, D. F., & Duit, R. (2015). On the significance of conceptual metaphors in teaching and learning science: Commentary on lancor; niebert and gropengiesser; and fuchs. *International Journal of Science Education*, *37*(5-6), 958–965.

Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in erp data using mixed-effects regression with R examples. *Psychophysiology*, *52*(1), 124–139.

Ubell, R. (2017). How the pioneers of the mooc got it wrong. *IEEE Spectrum*. (Accessed: 2018-06-28)

Ullén, F., de Manzano, Ö., Almeida, R., Magnusson, P. K., Pedersen, N. L., Nakamura, J., … Madison, G. (2012). Proneness for psychological flow in everyday life: Associations with personality and intelligence. *Personality and Individual Differences*, *52*(2), 167–172.

Vance, A., Jenkins, J. L., Anderson, B. B., Bjornn, D. K., & Kirwan, C. B. (2018). Tuning out security warnings: A longitudinal examination of habituation through fmri, eye tracking and field experiments. *MIS Quarterly*, *42*(2), 355–380.

Vance, A., Jenkins, J. L., Anderson, B. B., Kirwan, C. B., & Bjornn, D. (2019). Improving security behavior through better security message comprehension: fmri and eye-tracking insights. In *Information systems and neuroscience* (pp. 11–17). Springer.

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, *13*(2), 22.

Wagenmakers, E.-J., & Farrell, S. (2004). Aic model selection using akaike weights. *Psychonomic Bulletin & Review*, *11*(1), 192–196.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. *Nature*, *230*, 380–384.

Wammes, J. D., Boucher, P. O., Seli, P., Cheyne, J. A., & Smilek, D. (2016). Mind wandering during lectures i: Changes in rates across an entire semester. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 13.

Wammes, J. D., Seli, P., Cheyne, J. A., Boucher, P. O., & Smilek, D. (2016). Mind wandering during lectures ii: Relation to academic performance. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 33.

Wammes, J. D., Seli, P., & Smilek, D. (2018). Mind-wandering in educational settings. *The Oxford Handbook of Spontaneous Thought: Mind-Wandering, Creativity, and Dreaming*.

Wammes, J. D., & Smilek, D. (2017). Examining the influence of lecture format on degree of mind wandering. *Journal of Applied Research in Memory and Cognition*, *6*(2), 174–184.

Wickens, C. D., Gordon, S. E., Liu, Y., & Lee, J. D. (1998). *An introduction to human factors engineering.* Longman New York.

Young, M. S., Robinson, S., & Alberts, P. (2009). Students pay attention!: Combating the vigilance decrement to improve learning during lectures. *Active Learning in Higher Education*, *10*(1), 41-55. doi: 10.1177/1469787408100194

Zhang, J. (2016). Can MOOCs be interesting to students? an experimental investigation from regulatory focus perspective. *Computers & Education*, *95*, 340–351.

Zhang, Y., & Kumada, T. (2017). Relationship between workload and mind-wandering in simulated driving. *PloS One*, *12*(5).

Zhou, M. (2016). Chinese university students' acceptance of MOOCs: A self-determination perspective. *Computers & Education*, *92*, 194–203.

Appendices

Appendix A

Multiple Choice Quiz from Experiment 1

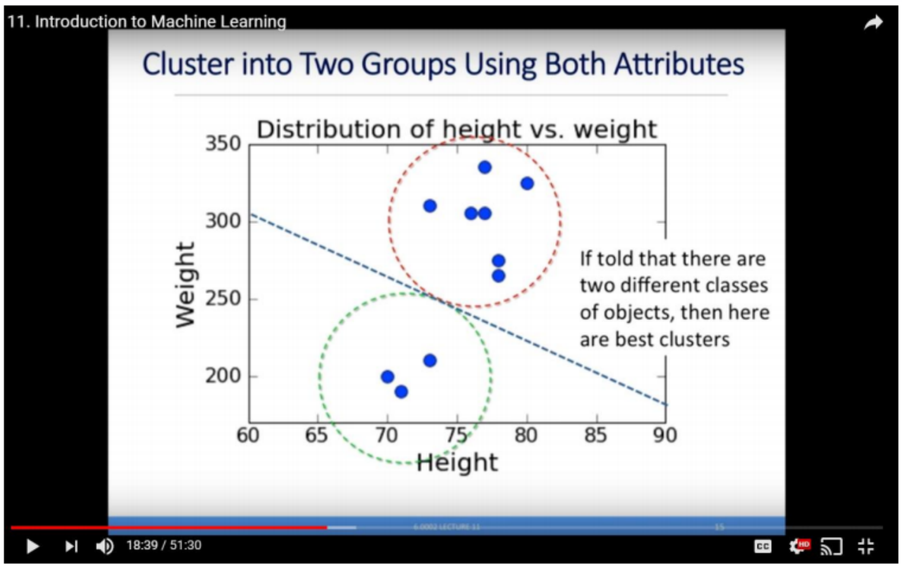
1. Which one of the following is an example of an application of a machine learning algorithm?
   (a) Netflix video rendering process.
   (b) *Two Sigma hedge fund predictions.*
   (c) Google Go programming language.
   (d) All of the above.

2. Of the following definitions, which best describes machine learning?
   (a) *A field of study that gives computers the ability to learn without being explicitly programmed.*
   (b) A field of study that gives computers the ability to learn by being explicitly programmed.
   (c) A field of study concerning the expression and mechanization of algorithms that underline the processing of information.
   (d) A field of study concerning the creation and implementation of algorithms that underline the processing of big data.

3. How is training data used in a machine learning task?
   (a) Given a set of feature/label pairs, training data is used to predict the label associated with the previously unseen input.
   (b) Given a set of feature vectors (without labels) training data is used to create natural clusters (or create labels for groups).
   (c) *Training data describes a set of examples used to infer something about a process.*

(d) Training data describes a set of previously unseen examples used to improve the algorithm.

4. The following image describes an aspect of machine learning. Which of the following best describes this aspect?



(a) Training data.

(b) Test data.

(c) Supervised learning.

(d) *Unsupervised learning.*

5. Grimson describes the concept of feature engineering. If we were to create an algorithm to predict student performance in a machine learning course, which of the following features would most probably cause the model to overfit?

(a) SAT scores.

(b) *Birth month.*

(c) Prior programming experience.

(d) Whether the student participates in the university debate club.

6. Which of the following best describes the concept of supervised learning?

   (a) Supervised learning uses clustering to group similar data into clusters.

   (b) Supervised learning uses manually created weights to measure Euclidean distance.

   (c) Supervised learning is a set of algorithms designed to prevent overfitting using the data features provided.

   (d) *Supervised learning uses labelled training data to infer a function.*

7. There is a picture below. Which of the following statements best describe the picture?

	Positive	Negative
Positive	12	0
Negative	9	9

   (a) *A confusion matrix.*

   (b) Data vectors.

   (c) Correctly labelled results.

   (d) Test data.

8. Are you paying attention to this quiz?

   (a) No.

   (b) *yes.*

   (c) No.

   (d) No.

9. Which is the best measure of a successful machine learning model?

   (a) Good test data.

   (b) High precision.

   (c) High accuracy.

   (d) *None of the above.*

Appendix B

Multiple Choice Quiz from Experiment 2

1. What is an example of an effective way that a Venture Capital partner can find investment opportunities?
   (a) To regularly read TechCrunch and identify promising start-up companies.
   (b) To do comprehensive industry analysis about technology trends.
   (c) *Host dinners and events to form a personal network.*
   (d) None of the above.

2. According to Ernestine Fu, what are the most important considerations in the VCs term sheets?
   (a) *Economics and Control.*
   (b) The pre-money and post-money valuations of the company.
   (c) The start-up team composition.
   (d) Liquidation preference.

3. The board of directors for a start-up company typically consists of:
   (a) Two start-up founders and the lead investor.
   (b) Two investors and the start-up CEO.
   (c) The start-up CEO, start-up CFO and an independent board member.
   (d) *The start-up CEO, the lead investor and an independent board member.*

4. According to Fu, what is the most likely way to become a VC partner?
   (a) To apply to a job posting while having 1-4 years management experience.
   (b) *To found a very successful company or to work as an executive at a successful start-up.*

(c) To be an angel investor who is constantly closing on start-up funding deals.

(d) To build a media presence and reputation in the valley before meeting other partners.

5. The following is an example of a common protective provision in a terms document:

   (a) The right for investors to withdraw their money in the firm at any time.

   (b) *An agreement that the next round of financing will be higher than the previous round.*

   (c) Insurance against employees taking back control over the firm.

   (d) Protection against other venture firms from investing in the firm

6. Why do venture capital firms insist on rights of first refusal?

   (a) *Due to SEC rules which regulate the maximum number of shareholders private companies can have.*

   (b) So that their start-up companies can prevent hostile takeovers.

   (c) To prevent employees from taking back control over the firm.

   (d) To prevent companies from pivoting their product direction.

7. What does pro rata financing mean?

   (a) *Maintaining percentage invested though future investment rounds.*

   (b) Financing that can be renegotiated at any time.

   (c) Financing that can be given at any time in the future.

   (d) Maintaining control in the companys board of directors in future financing rounds.

8. How much of the companys stock is typically reserved for the Employee Option Pool?

   (a) Less than 1%

   (b) 1-5%

(c) 5-10%

(d) 15-20%

9. Why do VCs insist on a No Shop Agreement when financing a start-up?

   (a) They have a fiduciary duty to their stakeholders.

   (b) Term sheets are not binding contracts.

   (c) *Both a) and b)*

   (d) Neither a) or b)

10. According to Fu, VCs are more likely to spend time with portfolio companies that:

    (a) Are struggling and could benefit most from the attention.

    (b) Are the performing the best.

    (c) *Have the greatest potential to perform.*

    (d) Have the highest valuation.

Appendix C

Post Study Questionnaire Instrument

On a scale of 1 (lowest) and 7 (highest) please answer the following questions about your experience throughout the lecture. Designation in brackets indicates the construct and source of the item.

1. I was interested in the video content. [control item]
2. I focused my total attention on understanding the lecture content. [NASA TLX]
3. I thought about strategies for staying focused on the content. [NASA TLX]
4. I thought ahead about what I would need to do to understand the lecture content. [NASA TLX]
5. I focused my attention on correctly understanding as much as I could from the lecture. [NASA TLX]
6. The pace of the lecture felt hurried or rushed. [NASA TLX]
7. I was successful at learning the content of the video. [NASA TLX]
8. I had to work hard to learn the video content. [NASA TLX]
9. I was discouraged, irritated, stressed or annoyed while learning the video content. [NASA TLX]
10. Learning the video content was mentally demanding. [NASA TLX]
11. Learning the video content was physically demanding. [NASA TLX]
12. I thought about members of my family. [Technology-Unrelated Mind Wandering]
13. I thought about friends. [Technology-Unrelated Mind Wandering]

14. I thought about something that made me feel guilty. [Technology-Unrelated Mind Wandering]
15. I thought about personal worries. [Technology-Unrelated Mind Wandering]
16. I thought about something that made me feel angry. [Technology-Unrelated Mind Wandering]
17. I thought about something that happened earlier today. [Technology-Unrelated Mind Wandering]
18. I thought about something that happened in the recent past (last few days, but not today). [Technology-Unrelated Mind Wandering]
19. I thought about something that happened in the distant past. [Technology-Unrelated Mind Wandering]
20. I thought about something that might happen in the future. [Technology-Unrelated Mind Wandering]
21. I thought about checking my email. [Technology-Related Mind Wandering]
22. I thought about checking my social media (e.g. Facebook). [Technology-Related Mind Wandering]
23. I thought about browsing other stuff. [Technology-Related Mind Wandering]
24. I thought about checking my phone. [Technology-Related Mind Wandering]
25. I thought about doing other online activities (e.g. online shopping, online game). [Technology-Related Mind Wandering]

Appendix D

Experience Sampling Methods

Participants are briefed on these tasks before taking part in the respective experiment. Method 1 was the experience sampling method for the experiment described in Chapter 4, while Method 2 was the experience sampling method for the experiment described in Chapter 5.

1. Method 1 - Mind wandering describes self-generated thoughts about something other than the content of the video. For example, these can be about something that you experienced earlier today, about your friends, or even about the funny way something is drawn in the video. When you experience your mind wandering away from the task of learning from this video, please push the space bar.

2. Method 2 - Which of the following responses best characterizes your mental state JUST BEFORE this screen appeared?

   (a) Completely on task

   (b) Mostly on task

   (c) Both on task and mind wandering

   (d) Mostly mind wandering

   (e) Completely mind wandering