Understanding Mental Workload in Everyday Life and its Role in the Future of Personal Informatics

Thesis submitted to the University of Nottingham for the degree of
Doctor of Philosophy
August 2022

Serena Midha

Supervised by

Max L. Wilson
Sarah Sharples

Abstract

Tasks are increasingly becoming cognitively based instead of physically based, and managing multiple tasks at once is becoming commonplace. Tools that help people to manage their cognitive activity in their lives could therefore be highly valuable. Unlike physical activity trackers, however, it is not yet understood how cognitive activity could be tracked in daily life in order to provide people with meaningful reflections and useful goals, known as personal informatics. Mental workload is a promising concept in this respect, due to its performance defining qualities and ever-growing relevance. Thus, this thesis investigates mental workload tracking in everyday life.

Mental workload has typically been investigated from an isolated task-based, 'work'load perspective, predominantly in safety-critical environments, meaning that our understanding of how mental workload functions in daily life is limited. By adopting a novel longitudinal, holistic, and person-centred perspective, the research presented in this thesis aimed to improve understanding of 1) physiological mental workload measurements in real-world environments, 2) how mental workload could be useful as a form of personal informatics and mental workload as a concept itself, 3) how mental workload data can be meaningfully communicated to users, and 4) ethical considerations for tracking devices. Two empirical studies were conducted

in relation to this.

Firstly, a naturalistic laboratory study used brain imaging methods to physiologically measure mental workload levels for general work tasks. Office-worker participants completed personalised reading and writing tasks at different levels of difficulty. Verbal interruptions were incorporated and coffee drinking was largely unrestricted. Results found that the measure was sensitive to reading tasks but not writing tasks, which helped to identify challenges for real-world mental workload tracking in terms of maximising the sensitivity of the measures. Interruptions were also found to affect mental workload levels, and these findings were interpreted using mental workload models.

Study 2 had a quantitative and qualitative phase. It explored mental workload as as a concept and as a form of personal informatics. The quantitative phase involved participants subjectively tracking their mental workload levels at regular intervals; questionnaires related to wellbeing were also completed each evening. The qualitative phase interviewed the same participants in depth about their experiences and perceptions of mental workload and mental workload tracking devices. From this research, an apparent Mental Workload Cycle was developed, where participants aimed to fluctuate between low, medium, and high mental workload levels. This was because each level serves a purpose but sustaining any level for too long results in negative consequences. Factors were identified that could disrupt the Cycle, and indeed our quantitative data indicated that actual behaviour often did not align with qualitative preferences.

Qualitative insights also investigated the design of mental workload tracking technology in terms of design and ethical considerations. Design considerations related to metaphors, colours, shapes, and descriptors. Ethical

concerns related to data privacy, validity, misinterpretation, and personal identity.

The important questions for cognitive activity tracking and understanding mental workload in everyday life are human-computer interaction ones. These relate to, for example, what useful data consumer neurotechnology could be used to track, what goals we could set for healthy lives, and how personal cognitive informatics will relate to the pervasive way we use physical activity tracking. Towards understanding this future, this thesis makes four contributions. First, we identify challenges for physiologically measuring mental workload in uncontrolled environments. Second, we develop the Mental Workload Cycle, a model that progresses understanding of 1) how mental workload can be used for personal informatics, and 2) mental workload as a concept in terms of the factors that contribute of the states of overload and underload. Third, we produced design recommendations for communicating mental workload data. And fourth, we explicated ethical concerns for future consumer neurotechnology. These findings should be used to progress personal informatics and human factors research, and implicate the direction of consumer neurotechnology as it develops towards longitudinal tracking of cognitive activity.

Acknowledgements

I would like to thank the MRL and Horizon CDT for this opportunity and continuous support, especially from those in the office: Andrea, Laura, Monica, and Emma. Thank you to the EPSRC for financially supporting this project. A big thanks to Anna and Paul, who carefully examined this thesis.

My deepest gratitude goes to Max. Not only has he been the world's best supervisor, but he has also created a wonderful research group with an environment that gives everyone a voice and room to grow. I am also deeply grateful to Sarah, the other half of my supervision team. She has provided invaluable direction for this project, and I have felt very lucky to be guided by such an intelligent and inspirational woman.

I would also like to thank everyone in my cohort for sharing this crazy journey with me. Thank you to the people in my research group: Horia, Adrian, Aleks, Johann, Elizabeth, Abi, and Jeremie, who have been such fun to work with. A special thanks to Adrian, who has been a wonderful friend as well as work friend; and to Aleks (and her parrot), who provided guidance when I really needed it.

Most importantly, I would like to thank my parents for their endless love and support, and for always encouraging me to reach for the stars in whatever I do. A special thank you to Sam, who has been my rock and best friend throughout. And thank you to Laura, my best friend, who has no idea what this PhD is about, but has helped me more than she knows.

This thesis is dedicated to two people who recently passed away. To my grandma, Margaret, who was so proud of me. And to Dom, my cohort colleague and friend, who did not get to finish his thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

We now have the ability to objectively track our physical activity whilst fulfilling the activities of our daily lives. Indeed, sensors that monitor physical activity have arguably become ubiquitous. Whether users choose to track, for example, their heart rate, steps walked, or calories expended, wearable physical activity trackers can facilitate positive behaviour change [180; 47]. Due to substantial technological progress in brain imaging methods, the widespread ability to measure interpretable cognitive activity in real-world environments is on the horizon [16; 14]. A logical and exciting next step is to consider what improvements could be made to our lives and work if (or when) we could track our cognitive activity in our daily lives in a similar way to our physical activity.

Neurotechnology devices that track brain activity in a healthy population is a market growing in quantity, quality, and investment. It is expected that commercialising neurotechnologies will provide a tool for improving health and wellbeing, productivty, entertainment, and education [73; 139]. Over the past 20 years, over $19 billion has been invested into neurotechnology companies [56] and patents being filed are only increasing[1]. Many of these devices monitor cognitive or emotional states of a user from arbitrary brain activity, and are referred to as passive brain-computer interface (pBCI) devices [14; 245; 244]. The field is rapidly progressing, and pBCI devices are beginning to emerge on the consumer market. Currently available devices aim to help users, for example, focus[2] or meditate[3]. This type of tracking is over a short and designated period of time, aiming to improve the cognitive state of the user; this can be considered similarly to how we can track physical activity data for a specific workout. In this respect, choosing to do a meditation session and track it with the Muse pBCI device is like choosing to go for a run and track it with a watch. But life improvements can also be made by tracking physical activity longitudinally. Tracking and reflecting on personal data to provide actionable insights is know as personal informatics [135; 145; 82]. For example, we know that walking 10,000 steps each day is generally healthy for most people, and we can track, reflect, and act on that data over days, months, or even years. However, as a parallel to this, we do not know what type of cognitive activity might be useful to track longitudinally as a form of personal cognitive informatics. Indeed, as life increases in cognitive complexity, managing our cognitive activity in our daily lives seems increasingly important.

Mental workload is a defining factor for how well an individual is able to perform on a task. This is generally in terms of how well an individual's available cognitive resources are able to meet the demands of a task [240]. Mental workload is a substantial area of research in human factors, where

[1] https://sharpbrains.com/pervasive-neurotechnology/
[2] Neurosity - https://neurosity.co
[3] Muse2 - https://choosemuse.com/muse-2/

it is predominantly investigated in safety-critical contexts in order to prevent overload and underload, which are the states where mistakes are likely to happen [240; 204]. The state of overload occurs when the demands of a task exceed the available cognitive resources [132], and underload occurs when there is too little stimulation [241]; these states can result in poor performance through errors and lapses in attention [235]. Thus, if a pilot, driver, or air traffic controller, for example, becomes overloaded or underloaded during their work tasks, serious incidents can happen [240]. Because of these implications, a large and increasing body of research seeks to physiologically measure mental workload in real-world work tasks, ultimately aiming to improve safety at work. Progress in this area has been significant, and it is speculated that brain imaging methods will enable the objective quantification of real-world mental workload levels [240].

However, because of its performance defining qualities, mental workload is also a meaningful concept outside of safety-critical work tasks. In the context of broader work tasks, such as a lawyer completing important paperwork, costly performance errors can happen if the worker becomes overloaded or underloaded. In this regard, mental workload is also a critical concept for successful task completion in the context of life activities outside of work, such as using navigation systems [203]. As well as its significance to daily tasks, the world is becoming more technological, meaning that tasks are increasingly being characterised by their cognitive demands rather than their physical ones [240; 204]. The work-life balance is also becoming ever more blurred [77], and we have a tendency to try and optimise our productivity and performance in our lives [53]. It therefore seems prudent to recognise that the concept of mental workload has never been so relevant to our daily lives. Thus, mental workload appears to be a strong candidate for a type of cognitive activity that may be useful to track as a form of personal cognitive informatics [236].

In this sense, the role that mental workload plays in our daily lives is not yet clear, meaning that our understanding of how we can track this data to make life improvements is limited. Currently, mental workload as a concept has been considered and measured from an isolated task perspective, in terms of quantifying the moment that an individual is becoming overload or underloaded [204; 240]. This is reflected in mental workload theories and measurement tools that have typically been used for the duration of work tasks. It is notable that a fundamental aspect of the mental workload concept is an individual's internal response ability, yet despite the emphasis on this person-specific nature, research continues to focus on mental workload from a 'numbers perspective' without a qualitative or holistic insight into how individuals personally perceive and approach mental workload. Therefore, the implications of approaching mental workload from a holistic, person-centred and daily life perspective could be two-fold.

Firstly, by developing a life view of mental workload and the impact it may have on our lives (and vice verse), we could start to understand how mental workload data could be useful as a form of personal informatics, and what goals we should be setting in that regard. This could be useful for future pBCIs as they develop towards more longitudinal tracking. Secondly, an improved understanding of mental workload as a concept could be developed by broadly considering the factors that could contribute to overload and underload, which are the states that mental workload research generally focusses on preventing.

1.2 Aims and Research Questions

The perspective outlined above equates to our motivation to measure mental workload in daily life. There is one overarching research question for this thesis, involving four different sub-questions, outlined below. The specific aims and aspects of of each sub-question will then be further outlined in turn.

- How and why should we track mental workload in everyday life?

  (a) Can we physiologically track mental workload levels in general work tasks, and what are the practical concerns of doing that?

  (b) Can a longitudinal and holistic approach to mental workload improve understanding of how mental workload could be valuable as a form of personal informatics, and mental workload as a concept itself?

  (c) How can objective mental workload tracking data be meaningfully communicated to users?

  (d) What should be ethically considered when developing mental workload pBCI devices, or neurotechnology in general?

RQ (a) As mentioned, objective measurements of mental workload are predominantly investigated and implemented in the context of safety-critical tasks [240; 204]. Therefore, a first step towards daily life measurements of mental workload is to investigate how we can measure mental workload for more general tasks. This is in terms of investigating the sensitivity of physiological measures (for example, can we differentiate between easy emails and hard emails?) and the impact on these measures from stepping away from controlled safety-critical environments (for example, do social interruptions confound measurements?).

RQ (b) As described above, a longitudinal, holistic, and person-centred approach to mental workload is different to the approaches that have previously been adopted. This approach could thus increase understanding of mental workload from a personal cognitive informatics perspective, which could guide the development of future pBCI devices. In this respect, in order to address 'how' mental workload data could be useful to track, we need to understand mental workload from a broader life perspective in terms of how it is perceived, how it is approached, and the impact it has on people's lives. Additionally, research from this novel perspective is also an opportunity for new insights into mental workload as a concept itself.

RQs (c) and (d) These research aims relate to the interaction aspects between pBCI devices and users, in terms of creating meaningful and safe interactions. Firstly, cognitive activity data is complex in nature, and thus it is also important to consider how the data can be meaningfully communicated to users. Indeed, effectively communicating data from personal informatics tools is vital for positive interactions from users that enable meaningful reflections and long-term use for supporting the optimisation of certain behaviours [84; 145; 120; 49; 50; 59]. Secondly, there is huge potential for unintended consequences that may arise with the development of consumer pBCIs devices. Therefore, it is essential to consider the ethical implications of neurotechnology as developments continue to progress [139; 73; 228; 125; 216].

1.3 Thesis Contributions

This thesis is multidisciplinary in nature and contributes to the fields of pBCIs, personal informatics, and human factors, from an overarching HCI perspective that seeks to create systems that are intuitive and usable and that seamlessly align with users' intentions.

More specifically, the pBCI field seeks to develop neurotechnology for the purpose of improving the cognitive states of users, where one focus is mental workload measurement for work tasks [14]. The field of personal informatics aims to develop systems that help people collect personal data for insightful and actionable reflection [145]. And the field of human factors aims to design systems around human's capabilities [204], aiming to optimise efficiency and performance, and largely focusses on mental workload management.

From conducting two empirical studies, the contributions from this thesis can be summarised as having:

1. Identified challenges for measuring mental workload across general work tasks.
2. Developed a new model of mental workload, specifically the Mental Workload Cycle, which improves our understanding of mental workload and how that data can be used as personal cognitive informatics.
3. Produced design recommendations for mental workload data.
4. Explicated ethical concerns for the development of pBCIs.

1.4 Thesis Overview

Table 1.1 summarises each chapter of the thesis. The table outlines the topic of each chapter and a brief summary of its contents.

Chapter 2 outlines related work that forms the foundations of this thesis. This is in terms of mental workload, personal informatics, brain-computer interfaces, and neuroethics literature.

Chapter 3 presents the first empirical study, which addresses research question (a) and investigates what it means to physiologically measure mental workload variations for general work tasks.

Chapter 4 introduces the second study, which has a quantitative and qualitative phase.

Chapter 5 presents findings from the second study that relate to using mental workload as personal cognitive informatics. It addresses research question (b).

Chapter 6 addresses research question (c) and presents findings from the qualitative phase of Study 2 regarding the design of mental workload personal cognitive informatics tools.

Chapter 7 addresses research question (d) and presents results from the qualitative phase regarding the ethical considerations of pBCI devices.

Chapters 8 and 9 then close the thesis with general discussions and conclusions.

It is important to note that the parameters of this thesis concern the life improvements that could be made through tracking mental workload data in terms of *wellbeing*, and not mental health. Whilst these topics may be of interest to mental health research, this remains outside of the scope of this project.

Table 1.1: Thesis overview table

Chapter	**Topic**	**Description**
Chapter 2	Overview of mental workload, personal informatics, BCIs, and neuroethics	Literature Review
Chapter 3	Physiologically tracking mental workload for general tasks	-Research question (a) -Empirical study -fNIRS to measure mental workload for reading and writing tasks -Published in: International Journal of Human-Computer Studies
Chapter 4	Outline of study 2 phases	-Design of a quantitative and qualitative empirical study
Chapter 5	Mental workload data for personal cognitive informatics	-Research question (b) -Qualitative and quantitative insights into personal experiences of mental workload -Published in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)
Chapter 6	Visual perceptions of mental workload trackers	-Research question (c) -Data from the qualitative phase of study 2 -Qualitative insights relating to design perceptions
Chapter 7	Ethical perceptions of pBCI devices	-Research question (d) -Data from the qualitative phase of study 2 -Qualitative insights relating to ethical considerations -Published in: 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)
Chapter 8	Discussions	Interpretations and future directions
Chapter 9	Conclusions	Final thesis conclusions

Chapter 2

Literature Review

2.1 Mental Workload

2.1.1 Characteristics

Mental workload is a meaningful concept in our everyday lives. It is a concept that can be interpreted intuitively [240] resonating with specialists as well as laypersons [184]; it can also be adapted to fit into many different contexts [184]. Our world is becoming more technology focused, and thus mental workload is a topic of increasing importance [240] due to tasks being characterised by their cognitive demands rather than their physical demands [204; 240]. If the mental workload level required to complete a task can be measured and managed, performance of the task may be successful, especially as the individual(s) is able to respond to unexpected situations [159]. For example, a pilot could complete a journey safely despite challenging weather conditions.

Despite discussions and research into mental workload being long established, numerous, and ever growing, there is no definition of mental workload that is widely accepted or used. In 1979, Moray [162] collected numerous papers by different contributors devoted to developing a universally agreed upon definition of mental workload. The significant and wide array of definitions that were proposed can help us to appreciate the difficulty of developing a definition of mental workload [204].

However, there is common ground between the different interpretations of mental workload. The components of mental workload are generally agreed to comprise of a) the demands of the task and b) the experience of responding to the task [203], in terms of the resources available to meet the demands [240].

The demands of the task reflect the characteristics of the task itself. This can have multiple facets, including the complexity of the task and time pressure for completion [240]. Further, Sharples and Megaw [204] outline how the externally measurable demands of the task may differ to the demand perceived by the individual, emphasising the importance of investigating how the demand is perceived by the individual.

The component relating to the experience of responding to the task regards how much ‘strain’ the individual performing the task is under [204; 240]. Young et al [240] outline that the effect on the individual is in relation the amount of resources available; these resources generally refer to attentional resources [230; 132], due to its critical role in the accurate and efficient processing of information [108].

To a degree, environmental factors can affect both the demands of the task and the experience of responding to the task [204; 240]. These can be external, such as other team members available to help with the task [204]. They can also be internal, such as skill level; in this case, increased skill level when it comes to the task will result in increased automatic processing [200], meaning attentional resources will be spared and mental workload levels will be lower compared to a situation in which the individual had lower skill levels [240].

Hence, mental workload is widely considered in terms of the resources available to meet the demands [227; 240]. These components that form mental workload as a concept critically affect task performance, as described below.

Firstly, however, we wish to differentiate between the concepts of mental workload and stress. These concepts are related as they share an overlapping characteristic in terms of a person's experience of responding to a task [8]. If an individual appraises a task to be beyond their capabilities, this could result in feelings of stress as well as high mental workload levels. However, whilst people often find high mental workload tasks to be stressful [8], stress as a concept is defined as an emotional state that is generated by the evaluation a person makes about their environment; a threat to the person's wellbeing will be perceived if they do not believe that they have enough social or personal resources to cope with the task [142; 239; 167]. Thus, mental workload is different to stress as it is not considered to be an emotional state, but is considered as the cognitive ability of an individual to perform on a task, in terms of amount of resources available to cope with the demands. This means that high mental workload states can be experienced without feelings of emotional stress if the task is not negatively appraised by the individual. The relationship between mental workload and stress will be explored in Chapter 5 in terms of the impact that daily mental workload levels can have on feelings of stress, but this thesis focusses on the concept of mental workload and not stress.

2.1.2 Mental Workload Models

Information Processing

Models relating to the concept of mental workload have been built upon this notion of limited resources [204]. In early work, including two well-known and similar models from Welford [227] and Whiting [229], this was described in terms of models of information processing. These models reflect how we process input that our senses have detected, similarly to how a computer functions in terms of receiving input, processing the information, storing new and retrieving old information, and generating an output. The concept of limited capacity was touched upon in these models by identifying that performance levels will drop if the demands of a task exceed the limited amount of information processing capacity that each human has. However, limited capacity was not well defined [204], and later models built upon and refined the concept.

In 1973, Kahneman [132] introduced the limited-capacity model of *attention*, which was fundamental for how we now consider resources in terms of attention, as described earlier. This model posits that an individual's attentional capacity at any one time is limited. Further, the total available capacity can be considered in terms of the capacity being devoted to the primary task, and spare capacity, which can be devoted elsewhere to, for example, a secondary task.

Wickens [233] later proposed an influential general model of information processing (Figure 2.1). Here, attentional resources need to be shared between different psychological processes involved in information processing (such as perception); as mentioned previously, the demands of the task influence the resources needed. Hence, mental workload can be considered in terms of the attentional resources distributed to these psychological processes and the demands of the task [204]. This leads onto the visualisation of the Limited Resource Model, originally outlined by Wickens et al in 2013 [232] and adapted by Sharples and Megaw in 2015 [204].

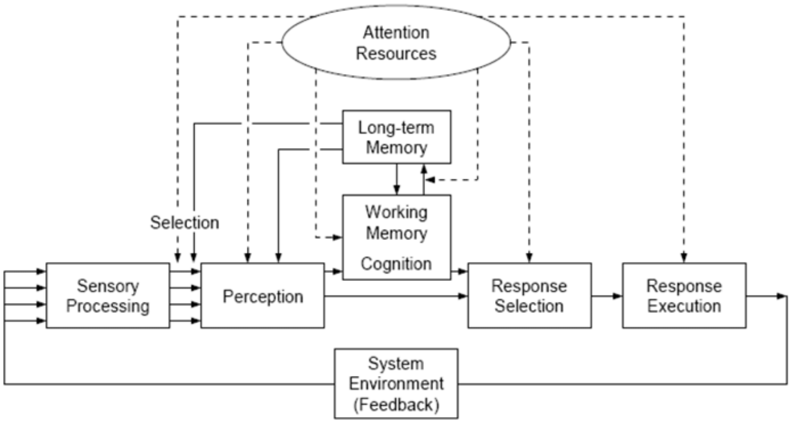


Figure 2.1: Model of information processing [233] from [204].

Limited Resource Model

The Limited Resource Model (Figure 2.2) displays mental workload as a product of the relationship between attentional resources and task demands.

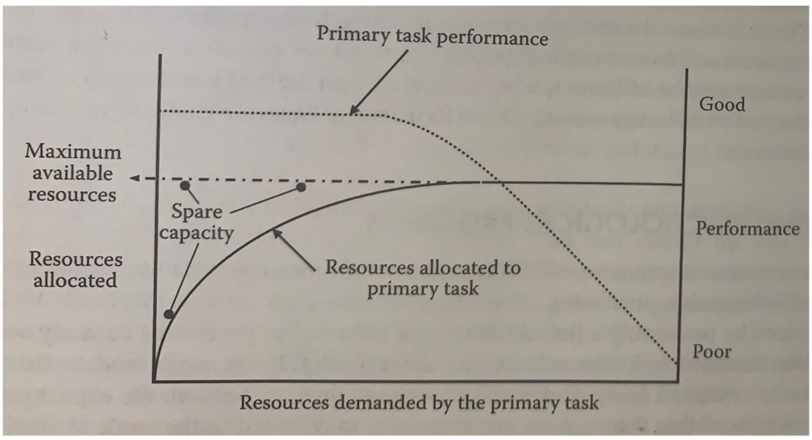


Figure 2.2: Limited Resource Model [232] from [204]

The vertical axis on the left show the amount of resources allocated to a task and the limit of resources that are available to the individual. The axis on the right indicates how well the individual is able to perform on the task. The horizontal axis shows the demands of the task. We can see by looking at the interaction of the lines denoting primary task performance and resources allocated to primary task, that if the resources demanded by the task exceed the amount of resources that the individual has available, the task performance will be poor. Likewise, task performance is theorised to be good if the resources demanded by the task are below the amount of resources that are available. Further, the difference between the resources allocated to the primary task and the maximum amount of resources available to the individual, can be considered as spare capacity.

When the demands of the task exceed the available resources, this is known as overload [132]. When there is too little stimulation, the available resources are either focused outside of the task or reduced because of underuse, and this is known as underload [241]. The states of overload and underload can both be very detrimental to performance, leading to attentional lapses and errors [235].

Sharples and Megaw [204] further developed the Limited Resource Model (Figure 2.3) to account for underload as a predictor of performance degradation as opposed to just overload. They also incorporated into the revised model how performance can drop due to data limitation, such that the information that needs processing is lacking in quality, such as the representation of a memory. In addition to this, the impact of overload on performance is not represented so 'gracefully', such that performance is likely to drop dramatically once the individual is overloaded. A final notable revision to the model, is the incorporation of varying levels of an individual's maximum resources, which can vary by both individual (e.g. cognitive span) and situational (e.g. alertness, vigilance, and fatigue) factors.

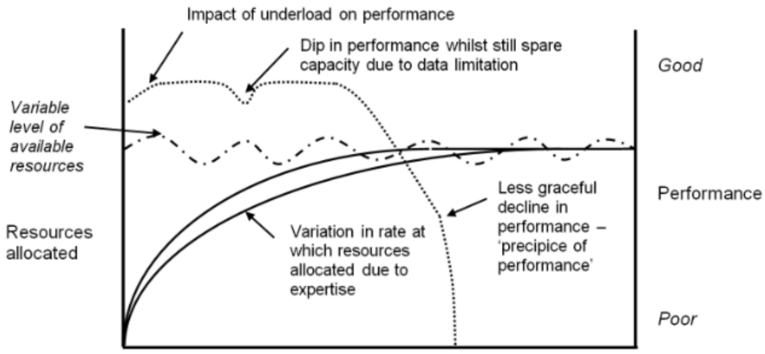


Figure 2.3: Adapted Limited Resource Model [204].

However, neither version of the Limited Resource Model accounts for the instances in which more than one task is performed at the same time. Secondary tasks, or subtasks, may require different allocations of attentional resources. The Multiple Resource Model [230; 231] describes the effect of multitasking on performance, and is described below.

Multiple Resource Model

The Multiple Resource Model [230; 231] (Figure 2.4) recognises that primary tasks are often make up of a number of tasks, and that more than one task is often performed at a time. Wickens described how tasks compete for a shared pool of multiple attentional resources. The theory has four dimensions. It begins with how the information for the task is processed (stages). Firstly, information is perceived, and this can be of an auditory or visual nature (modalities); information perceived visually (visual processing) can either be of a focal nature (such as reading text) or ambient nature (such as the perception and orientation of movement). Once perceived, the information is cognitively processed and then responded to. Information that has been perceived, processed, or responded to (at the stages of processing), can either be spatial or verbal (codes).

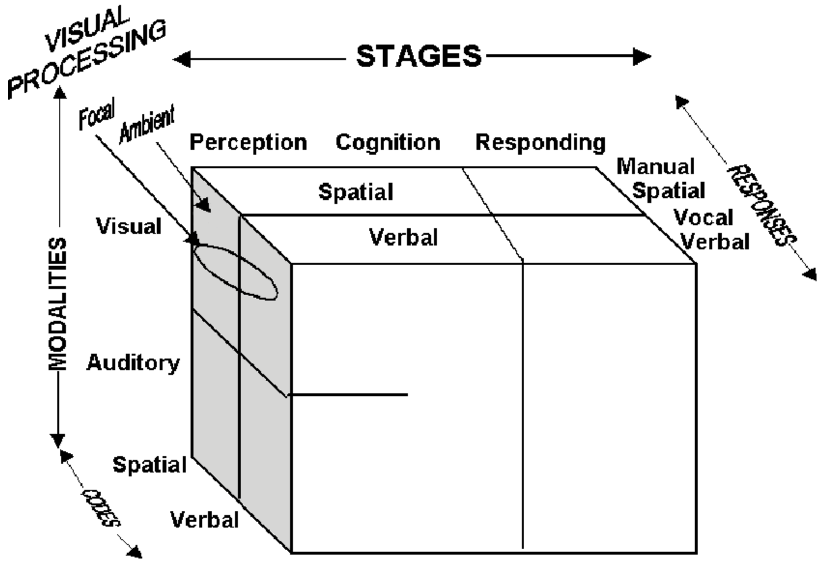


Figure 2.4: Multiple Resource Model [230; 231].

Tasks performed simultaneously may interfere with each other if they are competing for the same resources, but may be completed at the same time with good performance if they do not compete for or overlap in the resources that are required. For example, driving whilst texting can be disastrous because according to the model, they are both perceived visually, are spatially processed and require a manual response. Driving whilst having a conversation, however, is theoretically sound because different resources are required over the different dimensions.

Whilst the Multiple Resource Model is not designed exclusively as a theory for mental workload, they are closely related due to the theoretical explanation about how the demand of task(s) imposes on an individual's limited attentional resources, whether that be for isolated or multiple tasks at one time [231].

2.1.3 Measuring Mental Workload

As outlined above, mental workload is a multidimensional concept, and this is reflected in the numerous measurement techniques employed in applied settings.

O'Donnell and Eggemeier [172] outlined 5 criteria points that mental workload measurement methods should possess in order to measure mental workload levels in real-world environments robustly and efficiently. These included [172; 39]:

1) The method must be reliably sensitive to changes in task difficulty or resource demand and discriminate between significant variations in workload.

2) The method should be diagnostic, indicating the source of workload variation and quantify contribution by the type or resource demand.

3) The method should not be intrusive or interfere with performance of the operator's tasks, becoming a significant source of workload itself.

4) The method should be acceptable to the subjects, having face validity without being onerous.

5) The method should require minimal equipment that might impair the subject's performance.

Since then, Cain [39] outlined an expanded list of criteria to include:

6) The method should be timely and sufficiently rapid to apply to capture transient workload changes.

7) The method should be reliable, showing repeatability with small variance

compared with main effects.

8) The method should be selectively sensitive to differences in capacity demand and not to changes unrelated to mental workload (such as emotional stress.

9) The method should be insensitive to other task demands, such as physical activity beyond the conduct of the tasks [42; 39].

Mental workload measurement methods generally fall into three categories: performance measures, subjective measures, and physiological measures, which are outlined below.

Performance Measures

Performance measures of mental workload can be categorised into primary performance measures and secondary performance measures [108; 204; 39; 240]. The most widely used method of measuring mental workload is primary performance measures [240; 108], which involves directly evaluating an individual's performance on variables associated with the task. A common example for this is for driving, where efficiency of vehicle handling can be assessed from e.g. steering control or braking. If performance on these aspects is low, this could indicate that the demands of the task are too high and the individual is overloaded; in contrast, if performance on the task is sufficient, this could suggest that the demands of the task are within the individual's resource capabilities. However, this method used in isolation is limited. An individual might achieve high performance levels but to do this they need to input a lot of their available resources, such that their spare capacity is low; their performance level may indicate that their mental workload is at a comfortable level, but there are not enough resources to respond to an unexpected demands [204]. In addition, if performance is measured at short time intervals, individuals may naturally improve their performance when being monitored, or may not be able to maintain their level of performance for time after the performance measure has been taken [204]. Because of this disconnect between primary task performance and mental workload levels, it is advised to use more than one measure of mental workload [108].

Secondary task measures are used to provide an understanding of how much spare capacity an individual has whilst completing a primary task [39]. This is implemented by increasing the individual's mental workload to the point where they can no longer maintain their performance on the primary task [39]. However, this method violates the criteria for mental workload measures outlined by [172], as by nature it is intrusive and interferes with the task, such that it is limited to training environments and cannot be deployed in the real-world [204].

Subjective Measures

Subjective measures are another popular method for measuring mental workload levels as they are low-cost and easy to administer. Some researchers believe that subjective measures provide the most accurate reflection of mental workload levels [110]. Subjective measures are likely to reflect the processes underlying task performance, such as the effort expended and the available capacity [164; 204].

Subjective scales for the measurement of mental workload can either be multidimensional or unidimensional. Multidimensional scales capture ratings for the different dimensions of mental workload. It has been suggested that multidimensional scales are the most sensitive for measuring mental workload. However, they are time-consuming to complete [184]. Unidimensional scales consider mental workload as as one continuum, such that levels can be represented through one number. It is argued in the literature that unidimensional scales provide an better global representation of mental workload levels compared to multidimensional scales, whilst being less time-consuming to administer [113].

The NASA Task Load Index (NASA-TLX) [110] is the most widely known and applied multidemensional measure. It was developed by Hart and Staveland [110] comprises of six 21 point scales where a rating of 0 equates to 'Very Low' and 20 to 'Very High'. The scales include mental demand, physical demand, temporal demand, performance, effort and frustration. (Figure 2.5). NASA-TLX is considered more accessible than another well-known method called the Subjective Workload Assessment Technique (SWAT) [195] which is very time-consuming, often taking an hour to implement [204].

Unidimensional scales often use terminology specific to the industry for which they were developed [184; 204]. For example, the Cooper-Harper Scale [60] or the AFFTC [9] which were developed for use within the aircraft industry. The Instantaneous Self-Assessment (ISA) scale [36] was developed for a range of mental workload assessment contexts and is simple to use and deploy. Pickup et al [184] highlights how the ISA has been found to be a useful tool in real-world air traffic control work environments. The scale is from 1 (low mental workload) to 5 (high mental workload), and individuals give their perceived mental workload ratings at regular intervals (Figure 2.6).

Subjectively self-tracking symptoms and behaviours is frequently practiced in the health and wellbeing space by patients with chronic conditions

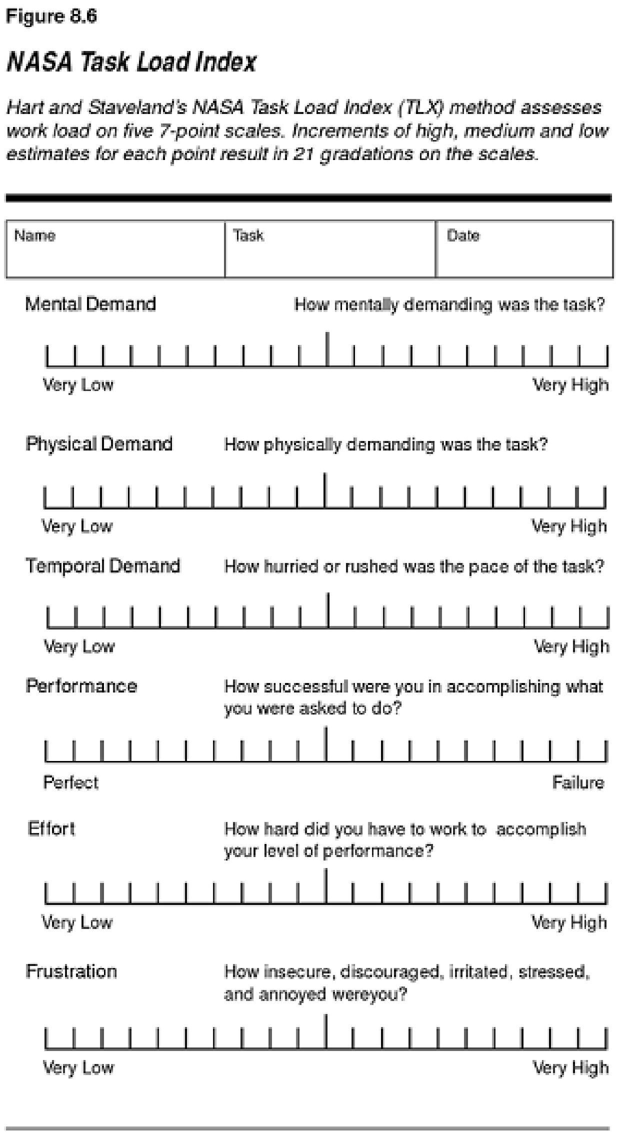


Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 2.5: NASA-TLX [110].

[170; 48; 24; 19]. This has been shown to increase patients' reflection, understanding of their condition, and support management behaviours [170]. However, manual self-tracking is inconvenient to complete, highly subjective, and often completed inconsistently [24]; the burden of self-tracking prevents people from adopting long-term self-tracking practices [82].

Further, subjective measures do not offer the potential for continuous and non-intrusive mental workload monitoring as they require input from the

Level	Workload	Spare Capacity	Description
1	**Under-utilised**	Very much	Little or nothing to do. Rather boring
2	**Relaxed**	Ample	More time than necessary to complete the tasks. Time passes slowly.
3	Comfortable	Some	The controller has enough work to keep him/her stimulated. All tasks are under control.
4	**High**	Very little	Certain non-essential tasks are postponed. Could not work at this level very long. Controller is working 'at the limit'. Time passes quickly.
5	Excessive	None	Some tasks are not completed. The controller is overloaded and does not feel in control.

Figure 2.6: ISA scale [36].

subject [204]. Ratings need to be given at task intervals, which may cause disruption to the task, or given retrospectively, which may be prone to inaccuracies. Referring to how subjective measures operate by measuring individuals' opinions, Gopher and Donchin [102] outline how "an operator is often an unreliable and invalid measuring instrument."

Physiological Measures

Because of the limitations of subjective measures, automatic monitoring devices are increasingly being researched and developed (such as wearable sensors), with the aim of lessening the burden of self-tracking whilst retaining its benefits [48]. Advances in technology, which are making substantial progress in the monitoring of certain cognitive and physical states, including mental workload, have the potential to enable long-term and continuous tracking.

Physiological measures stand on the assumption that as mental workload levels change, there will be a corresponding response in the autonomic nervous system which can be reflected and measured in a number of physiological parameters. Cardiac activity, respiration, electrodermal activity, eye function, and brain imaging methods are well-known physiological measures that have evidence supporting their use as tools for distinguishing between mental workload levels, and are outlined below.

Cardiac Activity

Charles and Nixon [45] outline how measures of heart rate are the most popular physiological method for measuring mental workload, measured through an electrocardiogram. This popularity especially applies to flight research [197]. Its wide spread use is likely because it is a practical method which is inexpensive and easy to use [119; 204]. Measures of cardiac activity generally refer to heart rate, characterising the number of beats over a period of time (typically a minute), and heart rate variability, which is the variation between each heartbeat in terms of time.

An increase in mental workload has been associated with a corresponding increase in heart rate [234; 197; 65]. An increase is mental workload has also been associated with a corresponding decrease in heart rate variability [219; 129; 108]. For example, Rivecourt et al [65] manipulated mental workload for pilots during a simulated flight task by varying the difficulty of certain maneuvers and found increased demand to be associated with increased heart rate and decreased heart rate variability.

Heart rate variability is generally considered to be a more sensitive measure than heart rate [1]. A relationship between heart rate and mental workload has not proven to be consistent [42; 107; 20], and this is perhaps because of the influence of physical activity on the measurement [204].

Respiration

Respiration is another physiological measure of mental workload. Respiration measures can refer to respiratory rate, airflow, volume, or respiratory gas analysis [45]. Respiratory rate is considered the most useful measure [197; 45], which tends to refer to the amount of breaths per minute.

For example, Backs et al [20] manipulated mental workload levels in an air traffic controller task by varying traffic volume and density across three conditions. They found that respiration rate significantly differed across each condition, where as the task demands increased there was a corresponding increase in respiration rate. However, measures of respiration are not very practical, especially for continuous monitoring of mental workload in everyday life. This is because respiration is heavily influenced by physical activity [103] (like cardiac measures). Also, speech production may affect measurements by interrupting and changing respiration patterns which impact respiration rate but are not related to changes in mental workload levels [45; 197].

Electrodermal Activity

Charles and Nixon [45] outline that electrodermal activity (EDA) refers to the change in electircal activity in the eccrine sweat glands. Because of this, EDA measures are affected by factors such as temperature, age, and sex which makes it difficult to compare between individuals or studies [138]. Still, Collet et al [57] measured EDA during real-world driving tasks and manipulated mental workload levels by varying the demand of braking. They found EDA to increase as task demands increased, indicating that EDA is sensitive to mental workload measurements.

Eye Function

Eye blinks, including of blink rate and blink duration, and pupil dilation are often used as physiological measures of mental workload. For example, Ricarte et al [65] found that higher blink rate was associated with a higher mental workload levels, generated by talking, listening, and calculating tasks. Ahlstrom and Friedman-Berg [4] found blink duration in air traffic controller participants to be significantly shorter during high mental workload tasks which included managing traffic in different weather conditions without the use of weather displays. Palinko et al [174] found an increase in pupil diameter during simulated driving tasks requiring increased mental workload levels.

However, in their recent review of physiological measurements of mental workload, Charles and Nixon [45] highlight that light quality, air quality, air conditioning, and drugs, can affect measures of eye function; thus, this method may not be practical for mental workload tracking in daily life, where these factors cannot be controlled.

Brain Imaging

Our understanding of cognition in the human brain has traditionally been developed by laboratory controlled studies using simple paradigms and stimuli [158; 141]. This work has formed, and continues to form, the fundamentals of many principles within cognitive neuroscience regarding cognitive processes and functional brain organisation [158]. However, a large body of recent research has made huge progress in studying brain activity in real-world environments [158; 141] due to developments in theories, signal processing techniques, computational power and brain mapping tools

[158]. Whilst at an early stage, it is believed that taking experiments out of the lab and into real-world environments may be a game changer for developing our understanding of brain activity, as the data is representative of the complexities and conditions of daily life [158; 141].

Brain imaging techniques as a measure of mental workload are growing rapidly in popularity [153; 3; 17; 186; 242]. The number of brain imaging studies measuring mental workload in realistic and real-world tasks is increasing dramatically, and they are considered to be part of the next generation of mental workload studies that may enable the objective, continuous and non-intrusive quantification of mental workload in real-world environments [240].

Traditionally, participant movement has interrupted the imaging signal, meaning that mental workload data was indistinguishable from movement artefacts. Functional Magnetic Resonance Imagine (fMRI) is perhaps the imaging gold-standard, but even small amounts of motion still make the signal indistinguishable [192]. Due to technological progress, some imaging methods are now available to enable the investigation of cognition in ecological settings [67]. Electroencephalography (EEG) is the most commonly used technique for measuring mental workload [32]. EEG measures brain activation directly by recording electrical neuronal activity from electrodes placed over the scalp.

There are four types of EEG patterns that can be recognised in a recording, including alpha, beta, theta, and delta waves. They are distinguished by their differences in wave frequency and amplitude. The theta and alpha bands in particular has been associated with changes in mental workload levels [12; 14; 204]. There is evidence that EEG is sensitive to mental workload measurements in laboratory environments, [28; 37; 38; 10; 224], for example the measurement of mental workload during traditional n-back tasks [224; 38]. There is also increasing evidence for the ability of EEG to measure mental workload in more naturalistic tasks, especially in aviation [12; 14]. For example, Borghini et al [31] found EEG effective in detecting mental workload levels in a flight simulation task for novices. Flumeri et al [91] also found EEG to be an effective measure of mental workload from their study involving professional air traffic controllers completing traffic management tasks at varying levels of difficulty. Further, Arico et al [14; 12] outline that EEG has been used to measure mental workload effectively in tasks relating to driving [32], surgery, and power plant control centres.

Because EEG records electrical activity in the brain directly (instead of from a secondary source), it has high temporal resolution, but has a relatively weak spatial resolution [115]. It is known to also be susceptible to artefacts [115; 189], which could be the most problematic barrier to for the its adoption in real-world environments.

fNIRS

In recent years, functional Near-Infrared Spectroscopy (fNIRS) has gained a lot of momentum [67] and has shown promise as a non-invasive and movement tolerable brain scanner [209; 152; 3; 185; 153; 242].

fNIRS uses near-infrared light to measure changes in blood oxygenation in the brain. Brain activity can be indirectly evaluated from this based on the concept of neurovasuclar coupling in which active brain regions require increased blood flow to meet enhanced energy demands. For understanding the effort involved in everyday work tasks, the prefrontal cortex (PFC) is one area of the brain often measured in mental workload studies [161; 97; 94; 21] as it is an area associated with executive functions required for the cognitive processes that mental workload is comprised of.

fNIRS has a relatively good spatial and temporal resolution [115] and is robust against motion artefacts [115; 189]. As a non-invasive, portable and movement tolerant brain imaging method, fNIRS is arguably the most effective tool for measuring mental workload *in-the-wild*.

Please refer to Appendix D for my written guide to fNIRS, outlining in detail how it works, how the data is processed, and how it can be analysed.

An accumulation of literature has found that fNIRS can reliably measure mental workload in a controlled laboratory setting for both traditional and more naturalistic tasks. For example, Ayaz et al [18] had experienced air traffic controller participants perform an n-back task followed by a simulated air traffic control task whilst wearing fNIRS. Mental workload during the n-back task was manipulated by changing the target item between 0-4 numbers previously; mental workload during the air traffic control task was manipulated by changing the number of aircrafts in each sector to either 6, 12 or 18 aircrafts. Results found that for both tasks, fNIRS detected increased PFC oxygenation in conjunction with increased task difficulty, supporting fNIRS as an objective measure of mental workload. Further, Maior et al [152] measured mental workload using fNIRS and asked participants to perform a verbal and a spatial mental workload task and then compared the oxygenation from the tasks to a rest condition; it was found that fNIRS could reliably detect mental workload for both the verbal and spatial conditions. Additionally, Fishburn et al [89] found oxygenation measured by fNIRS to increase linearly with memory load from an n-back task; the sensitivity of fNIRS to mental workload was so satisfactory that the authors suggested it as a viable alternative to fMRI. Further to this, fNIRS has also been used to differentiate mental workload levels in remotely operated vehicle operational tasks [72], and driving tasks [93].

There is research beginning to study mental workload using fNIRS in more uncontrolled environments. For example, fNIRS was effective in distinguishing low and high mental workload levels in pianists playing music pieces [242], computer programmers comprehending programming languages [165], and a table tennis player playing at two levels of difficulty [23]. However, research so far does not represent mental workload measurements during standard office work tasks or take into account the factors that could impair mental workload measurements in real-life office-work applications due to the increased complexity of natural settings [14].

2.1.4 Application Areas for Mental Workload Measurement

The ability to measure mental workload in the real-world is of extreme relevance to the neuroergonomics and human factors disciplines [17; 187; 240]. Mental workload is therefore predominantly researched within these fields, where a primary focus is preventing 'crashes' in performance in largely safety-critical tasks. As mentioned, mental workload is a defining factor for performance at work, as if the demands of a task exceed the resources available, performance errors can happen through overload [132]. Similarly, if there is too little stimulation, errors can also happen in the form of underload [241].

If for example a driver, pilot, signaller, or air traffic controller becomes overloaded or underloaded when interracting with a system, mistakes and serious incidents can happen [240]. Hence, typically such work tasks and even shift patterns are designed to remain within employee's capabilities

[203].

Neuroergonomics is the study of the human brain in relation to performance at work and in everyday settings [178]. A large research area within neuroergonomics is using brain imaging methods to measure mental workload [177; 179] with the aim of improving safety and performance at work and in life [176]. Indeed, mental workload research has predominantly been researched in regards to transportation applications over the last two decades, including driving, rail, and air traffic control research [240].

But mental workload is not only relevant to performance in safety-critical situations, but performance in broader work scenarios too. For example, a lawyer submitting important paperwork, a banker making an investment, or an online sales worker trying to close a deal, are examples of office-type work where if the task does not remain within the worker's mental workload capabilities, costly performance errors might happen.

Outside of work, mental workload also remains a meaningful concept, with strong relevance to areas such as using medical devices at home, navigating using technology [203] or performing work-like tasks at home [77].

Approaching Mental Workload Holistically

Mental workload has typically been considered from a short term, task-based, 'work'load perspective. This is reflected in mental workload literature which seeks to track mental workload levels only for the duration of the task of interest. It is also reflected in the mental workload theories and frameworks, which also consider mental workload in terms of isolated tasks.

However, this thesis argues that the experience of responding to a task makes up half of the identified (and agreed upon) components in mental workload, but research largely only considers the resources available to individuals without considering to whom those resources belong in terms of individual characteristics. It is therefore argued that approaching mental workload from a more holistic perspective should involve understanding people's perceptions of the concept itself. This is because individual perceptions of mental workload might affect how people respond to the use of their resources, which may in turn affect their performances. Thus, understanding mental workload as a person-centred, whole entity could implicate the ways in which different people approach mental workload in their lives, and could also implicate further understanding of overload and underload at work in terms of the personal factors that contribute to these given states. Therefore, where most mental workload research focuses on the avoidance of overload and underload to improve performance at work [240], our focus remains on the mental workload levels in between these extremes.

2.2 Personal (Cognitive) Informatics

We have experienced a revolution of self-tracking [135], where wearables and smart phones enable us to monitor aspects about ourselves that perhaps once would have been unimaginable. From physical activity data, such as heart rate, steps walked, and calories burned, to monitoring finances, locations visited, and menstrual cycles - self-tracking is now at our fingertips. Wearable devices for physical activity are so prevalent that they may be considered ubiquitous.

The Quantified Self movement[1] is an international community that have

[1]https://quantifiedself.com/

an interest in gaining "self knowledge through numbers", and comprise both users and makers of self-tracking technology. Personal informatics is a relatively recent scientific field that is growing in line with the expansion of self-tracking technologies. Personal informatics concerns the collection and reflection of personal data [135; 145; 82]. The term is often used interchangeably with quantified self, but instead tends to focus on providing meaningful data insights [145] to facilitate favourable changes in behaviour [135; 210].

Personal informatics research has primarily been researched and developed for the health and wellbeing space [79]. Physical activity tracking, chronic condition symptom tracking, mental health tracking, sleep tracking, and food tracking were identified as the top five most prominent application areas for personal informatics publications in a recent mapping review by Epstein et al [79]. Personal informatics for productivity improvement was identified as the second most prominent research and development area [79]. In this regard, several personal informatics tools have been developed with the aim of supporting users to increase their productivity levels; they tend to work by collecting data about number of hours spent working, the length of time spent on particular documents, and when they have been distracted by non-work online activities, such as social media [58]. Thus, this further shows that there is demand within the health and wellbeing and productivity improvement space in regards to personal informatics.

2.2.1 Models

There are two main models of personal informatics that describe how users interact with personal informatics tools. These are the stage-based model [145] and the lived-informatics model [82]. The models help to increase our understanding of the steps and behaviours users of personal informatics tools might take, which can be used to guide the design of personal informatics tools. The stage-based model was developed with the aim of understanding the problems that users of personal informatics systems might face; the lived-informatics model aimed to revise and develop the stage-based model after arguing that in practice people do not adhere to the strict division of stages that were proposed.

Stage-Based Model

Li et al [145] proposed the stage-based model of personal informatics which provides a common framework to describe, compare, and evaluate personal informatics systems (Figure 2.7). It was developed by conducting surveys and interviews with people who already collect and reflect on their own personal data. The model comprises of five stages. The preparation stage is where users decide what information type to collect based on their motivation(s) and the technology they will use to collect it, and hence this stage occurs before any personal information is collected. The barriers at this stage relate to decisions about which information to collect and the method of collecting it. The collection stage is where users actually collect their personal data. This can vary in frequency, such as several times each day, once each day, or continuously. Barriers to collection included the tool itself, for example not having access to a computer when needing to log data, and the user, such as forgetting to log data or not having the motivation to do so. The integration stage is next, where the collected data is made into a format appropriate for the user to reflect on by perhaps combining data sources and transforming the data for reflection. Barriers at this stage can arise from, for example, manually organising or transcribing the data. The reflection stage is where users reflect on their personal data, including viewing or interacting with data visualisations. This can be a short time after the data has been collected in order to gain an understanding of the current status, or over the longer term, to observe trends or patterns in the data. At this stage, there can be problems with regards to the ability to explore (through time constraints) or understand the data. The final stage is the action stage, where users use their reflections to inform their behaviour; the increased understanding about their behaviours helps the user to make modifications in this respect. The tools themselves can also make recommendations about behaviour change based on the data, but Li et al (and others [58] systems that lack this feature are barriers to behaviour change.

Li et al [146] later expanded the reflection section to differentiate between users in a 'discovery' or a 'maintenance' phase, which affects the type of information that they desire to interact with. The authors outlined that whilst some users might be using the personal informatics tool to inform behaviour change (the discovery phase), other users may have already met their goals through behaviour change and are now using the personal informatics tool to maintain their behaviour habits instead of seeking to change them (the maintenance phase).

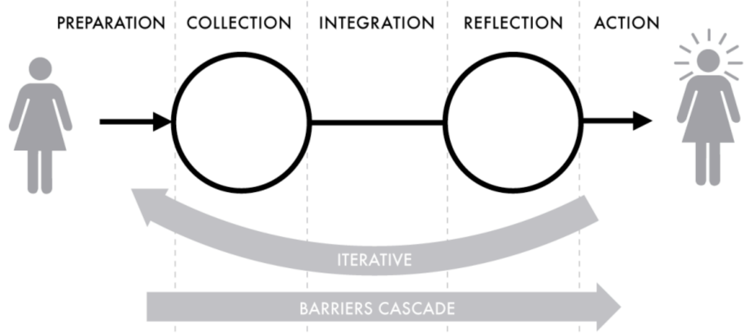


Figure 2.7: Stage-based model of personal informatics [145].

Lived-Informatics Model

Epstein et al [82] expanded upon the stage-based model, arguing that the stages do not capture the realities of self-tracking in practice. They proposed a model called the Lived Informatics Model (Figure 2.8) which was based on the notion of 'lived informatics', presented by Rooksby et al [196]. Lived informatics is a perspective in which it is appreciated how self-tracking technologies are a part of our everyday lives, and people engage with this data in different ways [196; 78]; this differs to the perspective presented in the stage-based model [145], which implies that data used for personal informatics has been thoroughly processed and analysed with goal setting in mind.

Rooksby et al [196] therefore identified five overlapping self-tracking styles: directive, documentary, diagnostic, collecting rewards, and fetishised. Directive tracking refers to users who track their data to help them to achieve a specific goal. Documentary tracking refers to users who wish to keep track of their activities but not change them. Diagnostic tracking refers to users who track their data in order to identify a relationship between two variables. Collecting rewards tracking identified users who tracked their data in order to score points or register achievements. Finally, fetishised trackers refer to users who track because of their interest in the technology and gadgets.

The lived-informatics model [82] was also developed through surveys and interviews of self-trackers. In this model, the preparation stage is divided into 'deciding' and 'selecting,' in terms of making the decision to track and deciding what to track, and selecting the method of tracking. Further, the stages of collection, reflection, and action from the stage-based model are integrated into one stage called 'tracking and acting.' This is because whilst the categories are distinct, they can occur simultaneously and are therefore considered together. At this stage, lapsing can occur, which is a temporary or permanent break in tracking. However, after a lapse, resuming tracking can occur and the user can go directly to the tracking and acting stage instead of returning to the deciding and selecting stage.

Whilst there are differences in the models of personal informatics, there are many similarities, and they both describe how personal data can lead to self-insight, and can facilitate positive behaviour change.

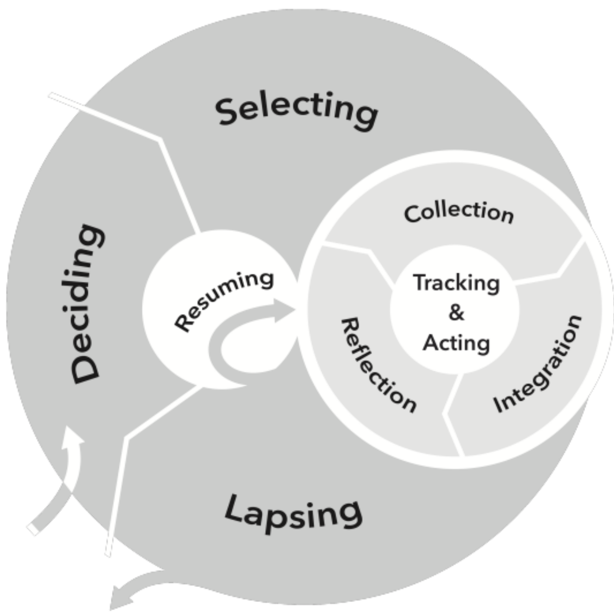


Figure 2.8: Lived Informatics Model of personal informatics [82].

2.2.2 Personal Cognitive Informatics

As described in Chapter 1, we are interested in exploring beyond short instances of cognitive activity tracking and researching cognitive activity tracking in daily life. Thus, there is a need to investigate what cognitive data would be useful for neurotechnology users to track longitudinally that enables goal setting for the optimisation of certain aspects of our lives. For some forms of personal informatics, users have found the data collected to not be useful for meaningful insights, which is a barrier to technology adoption [145] and continued use [80].

Further, Collins et al [58] identified four factors that were barriers for user engagement with productivity personal informatics tools, including salience (data should be easy to access), contextual information (output should support comparisons), credibility (users need to believe in the accuracy of the data), and action advice. Action advice refers to the tool providing information about what action the user could take based on the data, and this is particularly relevant for the ability to set goals. Ensuring that the tool supports actionable insights is not just relevant to productivity tools, but for personal informatics tools in general [145].

2.2.3 Mental Workload for Personal Cognitive Informatics

With the world becoming less physical and more technology focused, mental workload is as relevant now as ever as a concept to be considered in our daily lives. Our work-life balance is becoming ever more blurred, we frequently perform cognitively based tasks outside of the workplace [77], we strive to optimise our efficiency and performance at work, and we seek to lead healthy and happy lives [53]. Research has shown that high levels of mental workload at work play a role in accidents at home [75], and the impact of the weekend can affect accidents at work [74]. Similarly, demands for mental effort in our home lives can lead to poor performance at work [128]. We do not have a clear picture, however, of how people would try to manage mental workload in these different cases if they could measure their brain data as a form of personal informatics, or how workplaces might adapt to understanding mental workload from a broader perspective.

With the continuous measurement of mental workload a real possibility on the horizon, mental workload could be useful as a form of personal informatics, like a Fitbit for the brain [237]. Here we try to draw useful (although not infallible) parallels with physical activity tracking. The new consumer neurotechnology devices (outlined in Section 2.3.3) take brain measurements comparable to gyroscopic data in phones and watches. Their interpretation of this data into physical activity e.g. steps or swimming strokes, would be comparable to their ability to make inferences about relaxation or focused attention, or indeed mental workload levels. Beyond this, it is unclear what people would want to know, or indeed try to achieve, if they had mental workload as a form of personal informatics, as a parallel to trying to reach 10,000 steps a day, or train for improving fitness. It should be noted, however, that as mental workload is not a tangible concept like physical activity data, people are likely to be individual in the way that mental workload should be approached in their lives; this therefore emphasises the importance of a holistic approach.

2.2.4 Data Reflection and Visualisation

We know that people wish to optimise aspects of their lives, such as being fit and healthy, but there are often discrepancies between our desired and actual lifestyles that can sometimes be attributed to a series of poor and simple everyday decisions [59]. Research has been conducted towards identifying guidelines for the presentation of data that encourages positive behaviour change. For example, Consolvo et al [59] used psychological theories to outline eight design strategies for self-tracking data presentation, such as making the designs abstract, aesthetic, and positive. Choe et al [49] outlined design recommendations for the visualisation of personal data, including visualisations that encourage self-reflection, valid insights, effective data displays, and more playful visual annotations.

Moreover, several personal data visualisation displays have been developed based upon meaningful metaphors (e.g. [59; 84; 96]). For example, to encourage green transportation, an interface of a bare tree that grew leaves and fruit after every green transportation event was developed in addition to a polar bear standing on a small block of ice that got bigger with an improving surrounding ecosystem after each green transportation event [95]. Fish 'n' Steps [148] also aimed to display personal data meaningfully to encourage behaviour change by using a virtual fish in a tank that got bigger and happier as the users' step count increased, and remained smaller and angry or sad if the step count was not sufficient.

These personal data interfaces have been designed for use in a personal context as opposed to a professional context. The design focus centred on providing meaningful visualisations for non-expert visualisation or data analysts. This is an important area of research, as traditional visualisations aim to support expert analysts in their roles [120], but these styles of visualisations are not appropriate for lay people tracking personal data in a personal context [120; 50]. However, there has not been enough progress in data exploration and analytic capabilities for personal data analysis [86], and users are often left to do the heavy lifting [86].

Huang et al [120] noted that this may be because current designs are made by system designers who make decisions on how the data is presented without considering the unique perspectives of individuals. Hence, they called for further research aiming to explore how to support people so that they can increase their understanding of the personal data in which they choose to self-track [120]. In this vein, "data must be accessible, understandable, and interpretable before interacting with it can lead to insights or actionable knowledge" [120], and thus there is a need to investigate how users consider the aspects that they choose to self-track in order to develop effective ways of displaying the data back to them.

2.2.5 Designing for Cognitive Activity Trackers

It has been noted that personal informatics research has largely been focused on data collection rather than data presentation and interaction [120]. To establish what data matters to users, designers need to consider the ways in which people live, their concerns, and their internal conceptions [191]. Hence, as well as establishing which types of cognitive activity data might be useful, efficient, and healthy for users to track, there is a need to investigate effective ways of displaying the data back to users in order to maximise user engagement and long-term tracking, as outlined above.

Wilson et al [237] aimed to explore ways to visualise mental workload data so that users are able to gain valuable insights when reflecting on their data. By combining diary studies, interviews, and focus groups, their preliminary work started to paint a picture of example metaphors and descriptive words that people associate with different levels of mental workload. Displaying physical activity personal informatics data through metaphors has been previously researched, and compared to charts, they are found to be more engaging, motivating, glanceable, and ambient, whilst remaining effective at conveying information [84].

2.3 Brain-Computer Interfaces and HCI

Human-computer interaction (HCI) researchers continually seek to improve the quality and capabilities between humans and computers, ultimately aiming to create intuitive and usable systems that seamlessly fit with users' intentions. One way of achieving this has been by researching ways in which computers can communicate relevant information to users abundantly. Likewise, researching ways in which users can input information or commands into the system has also been a priority. In this sense, Tan and Nijholt [211] outlined how this input has generally required users to perform certain motor movements, such as moving and clicking a mouse. However, progress in brain imaging methods (as outlined earlier) has opened the door for enabling HCI research to progress to the next level. Prior to this, systems were evaluated purely by observable behavioural measures, such as performance measures or questionnaires [62]. Tan and Nijholt [211] described brain imaging methods as entering a maturity phase, where the concept of the ability to measure cognitive states has been proven and basic functions have been achieved, so designers can work towards using its unique attributes to create experiences that are novel and cannot be achieved with other technologies. Research in this sense is still at a relatively early stage, but, as mentioned, is progressing substantially and becoming closer to being implemented widely in real-world environments, and consumer neurotechnology (where products can be sold directly to consumers) is already starting to filter into the market on a small scale (outlined further in this section).

2.3.1 Brain-Computer Interfaces

The idea of Brain-Computer Interfaces (BCIs) was introduced in the early 1970s [221], where the original aim was to provide people with severe motor disabilities with means for communication and control [14]. In this sense, BCIs refer to the recording of brain signals, where features are extracted and converted into artificial outputs [14], such as mouse clicks on a computer. The field has expanded over the last decade, and now BCIs include applications relating to healthy subjects. Zander et al [245] therefore outlined three categories of BCI, consisting of active, reactive, and passive.

BCIs can be invasive or non-invasive [143]. Invasive BCIs refer to brain activity that is recorded from EEG electrodes that have been implanted into the brain. This means that the patient is required to undergo surgery in order to fit the system which carries surgical risk; the procedure is therefore usually only reserved for medical purposes, but does result in recorded signals of a high quality [118; 143]. Non-invasive BCIs regard brain activity that is recorded from the surface of the head, for example from EEG and fNIRS devices. This method can be applied on a wide scale to disabled an non-disabled people and has several areas of application. The recorded output is not as high in quality as invasive BCI [118; 143], but users have the advantage of not undergoing surgery, and the impact of non-invasive BCIs can be beneficial to many different areas [143], as well as potentially being flexible, quick, and easy to set up and use. Non-invasive BCI will be the focus of this section.

Active and Reactive BCI

Active BCI works towards enabling direct communication between a user and a technical system by using consciously controlled brain activity to control an application. Closely related is reactive BCI, which refers to how users control an application by their brain activity that occurs in response to externally presented stimuli.

Active and reactive BCIs tend to be considered as 'traditional' BCIs, which can increase independence of people suffering with conditions that disrupt their motor control, such as those with multiple sclerosis, [243] by bypassing the peripheral nervous system [238]. Application areas for these BCIs include communication purposes; for example, outlining simple 'yes' or 'no' decisions have been classified with 82 percent accuracy using fNIRS [166]. Motor assistance is another application, such as controlling a cursor to browse computer options and make a selection [52], and BCIs to aid movement is another application area, such as for the control of wheelchairs [212]. Additionally BCIs can be a valuable tool for stroke rehabilitation, for example by helping patients regain lost muscle function or coordination through motor imagery and neurofeedback that results in a change in neural connections [183].

Passive BCI

The notion of passive BCI (pBCI) was proposed by Zander et al [245] which aimed to capture all types of BCI that did not fall under the active or reactive categories. pBCIs refer to systems that monitor cognitive or emotional states identified through arbitrary brain activity [14; 245; 244]. Zander et al [245] outlined the key properties of pBCIs. Firstly, because of their ‘passive’ nature, they should be *complimentary*, meaning that it should not interfere with the technical system through direct control or communication like active or reactive BCIs do. pBCIs are also *compostable*, meaning multiple pBCI outputs can be recorded without conflicting with one another.

2.3.2 BCIs for HCI

Because of the development of BCIs, research in HCI has been able to progress to investigating how information or commands can be inputted into systems without being executed with motor movements, as well as how systems or tools can be evaluated objectively in real-time; these research areas have been made possible by progress in brain sensing, which can essentially detect a user’s thoughts.

Indeed, Nijholt et al [168] described three broad areas of HCI research that BCIs enable. The first is the ability to control computers with thought alone, which relate to the properties of active and reactive BCIs. This is particularly relevant for patients with motor disabilities, in which sophisticated systems can be developed to help increase their quality of life, as mentioned above. For people without physical disabilities, there is now the opportunity to research how to develop experiences of interacting with computer systems through thought, such as during gaming [168], and research in this area is developing [5].

The second and third areas of research are in relation to pBCIs. The second regards the evaluation of interfaces and systems [168], and this refers to how the cognitive or affective measures can be used to evaluate a system or user. Arico et al [14; 15] focussed on the user aspect of this research area and simply outlined that pBCIs could be used to provide feedback to users about their cognitive state. Nijholt et al [168] highlighted how this is particularly useful for evaluating mental workload levels either in terms of the users' capacity capabilities or the system's mental workload impact on the user in order to make improvements to the system. The use of brain imaging methods for monitoring mental workload levels in order to prevent overload and underload in mostly safety-critical situations has been outlined in section one, as measuring mental workload to provide feedback to users is an overlapping research aim in the neuroergonomics and BCI fields [14; 16]. This is speculated to be useful for improving work tasks [16], as mentioned previously.

Aside from mental workload research, which seemingly dominates research in safety operational contexts, especially in driving, rail, and aviation, other related factors have also been measured using pBCIs. In driving contexts, pBCIs have been used to measure attention levels on the task [225] , motion sickness whilst driving [147], and fatigue [223]. Fatigue in pilots has also been measured using pBCIs [67], and interestingly to monitor students' attention in lectures [137]; in that study, EEG was used in real-world lectures. Stimuli in the form of different shapes were presented on the screen, and attention levels were measured based on response times for students to notice the shapes.

On the other hand, Lukanov et al [151] provides an example of using pBCI to evaluate an interface. They used fNIRS to evaluate mental workload levels in response to filling out three different versions of a car insurance claim form. The versions varied by how the information was displayed to participants, such as the use of sub-forms. The authors found that fNIRS could provide insight into which layout generated the lowest amount of mental workload for participants, and suggested that this could be a way of reducing errors. Peck et al [182] also found fNIRS effective at detecting mental workload levels for participants interpreting bar charts and pie charts.

In addition, pBCIs are researched and used in terms of neuromarketing; this is because of their ability to detect feelings not overtly expressed by the user whilst interacting with marketing content, such as advertisements. [14]. As an example of this, Guixeres et al [105] used EEG to evalute the effectiveness of certain YouTube advertisements. The authors found that the EEG (and heart rate variability and eye tracking) measurements significantly correlated with metrics including advertisement recall, liking, and amount of views, suggesting that neuromarketing techniques could be useful for predicting the success of responses to advertisements. Similar results have consistently been reported elsewhere [41; 218; 46].

An extension of this second research area for pBCIs is the development of adaptive user interfaces [168], such that interfaces can adapt based on the cognitive state of the user. This is generally for the purpose of keeping users in an optimal state by managing factors such as mental workload, stress, and fatigue [14; 15; 240]. For applications relating to safety, mental workload levels were found be effectively managed in simulated air traffic controller tasks by measuring mental workload levels with EEG and adapting the interfaces (for example by highlighting specific information and filtering out non-critical information) to avoid overload or underload [12]. The authors note the potential of adaptive systems, but describe how there is a lack of examples in regards to mental workload, especially outside of laboratory settings. Additionally, in a real-world driving task, emergency braking systems could be effectively activated through passively monitoring cognition with EEG, and was found to save around 130ms of braking time [112].

In a learning context, Yuksel et al [242] developed an adapting system as an education tool; they increased the difficulty of piano pieces if fNIRS measurements indicated that participant's mental workload level was below a certain threshold by adding lines of music into the piece they were playing. Results found that participants could learn more accurately and could play faster music using this adaptive system compared to a control group that learned in a traditional way. Daley et al [63] studied how people's affective states (mood and emotions) can be controlled by using EEG (among other physiological measures) to infer the participant's emotional state and then playing targeted music to control their emotion. They found that they could effectively help participants to feel happier, calmer, and de-stressed.

Additionally, pBCI has been incorporated into gaming for entertainment purposes [157; 5; 14], by measuring the users' affective state and adapting specific features of the game accordingly, such as the difficulty level [12]. For example, van de Larr et al [217] incorporated pBCI into the popular game World of Warcraft. The user's character could take the form of either a wolf or a bear; the form of the elf enabled the user to attack enemies from a safe distance away, whereas the bear enabled close-up attacks. If the passive EEG measurements indicated that the player was in a relaxed state, their character would stay as a bear; if the player was assessed to be more agitated, they would become a bear.

2.3.3 Current Neurotechnology

Because of the substantial progress in the BCI field, it is believed that the technology is not far away from becoming freely available to consumers [16]. Devices are beginning to arrive, and a sample of currently available pBCI systems that are being sold directly to consumers are outlined below.

Neurosity

Neurosity [2] is a consumer EEG device that translates gamma activity into interpretable levels of focus. A baseline level for focus levels is created, and if gamma levels drop below a threshold, this is interpreted as a decrease in focus. If a decrease is detected, the Neurosity app plays music that has been specifically composed by the company to help the user concentrate; the choice of music is chosen depending on what it recognises that the user responds best to. The company outline that the result is a 'shift' in concentration that helps the user to increase their levels of concentration over 5x faster than without the use of their BCI device. The Neurosity app also enables users to keep track of their concentration levels in real-time or retrospectively.

Muse

Muse [3] also uses EEG for their pBCI device which aims to guide meditation sessions. This works by users connecting the device to the Muse app through their phones. When the system detects that the user is in a restful state, it will play calm music, such as peaceful weather sounds; when it detects that the user's focus is drifting, it plays more intense music, such as stormy weather sounds, in order to bring the user's attention back to their breathing. The app then provides feedback about the meditation session, using graphs to outline how much time was spent in active, neutral and calm meditative states.

[2] https://neurosity.co/
[3] https://choosemuse.com/muse-2/

Emotiv

Emotive [4] is a pBCI device which uses EEG to measure brain activity from each lobe of the brain. It claims to infer several cognitive states, and thus has several different uses. Emotiv can measure excitement, task engagement, relaxation, interest, stress, and focus levels. It appears that this data is passively measured and the information is relayed to users through an app; Emotiv does not appear to provide guidance or recommendations for how to improve their cognitive state.

2.3.4 Current Limitations of BCI development

Due to being at an early stage of development, consumer neurotechnology currently faces developmental issues. From a technological point of view, BCI technology is still making many misinterpretations about a user's cognitive state [14; 15; 118; 30]. However, as mentioned, research is making huge progress at overcoming barriers to the measurement of real-world brain activity, and the general consensus is that we are not far away from being able to accurately infer cognition in real-world environments. Additionally, neurotechnology outside of the medical field is largely legally unregulated [216], meaning users are not guaranteed that the data is valid and representative of true cognitive function [125; 73] (discussed more below). Current cognitive activity measures also need to be worn fairly obviously on the head, which may not currently be suitable for longitudinal use due to device discomfort and social aspects. However, it is likely that future technology will have the ability to track cognition from more commonly wearable sensors [8], such as the wrist or more subtly and comfortably from the head. For now, however, research is aiming to reduce the size and

[4] https://www.emotiv.com/

increase the ease of use of the brain scanners, but this can somewhat be at the expense of reducing the quality of the signal [14]. Arico et al [14] outlined a further issue for BCI devices, in that currently most pBCIs need calibrating frequently based on training data, which creates inconvenience for the user. In this regard, progress is being made towards reducing the frequency needed for this [14; 13; 226].

2.4 Neuroethics

As mentioned, the consumer neurotechnology market, which includes brain tracking devices, is rapidly growing in availability and investment [139]. But this growth in consumer neurotechnology comes with myriads of ethical considerations that must we must bear in mind [139; 73; 228; 125; 124; 216], and will be considered deeply in this thesis that seeks to progress the development of pBCI devices. Hence, neuroethics is a field that refers to a broad range of ethical, legal and social issues that have emerged through progress in neuroscience [85]. Giordano [100] describes how the public anticipates ethical issues incurred by the speed and breadth of neuroscientific discovery, and that whilst the future is full of possibilities for insights into our cognition, there is also potential for misuse of information, misunderstandings and foul play.

Specifically in terms of consumer neurotechnologies, Kreitmair [139] has outlined seven ethical dimensions that should be considered in the development through to consumption of these technologies. These regard how the products must firstly be safe, without any medical or cybersecurity risk. They should be transparent, meaning the products must be validated in their performance. Privacy is the third dimension, where consumer data should be handled responsibly, such that data remains private. The technologies should be epistemically appropriate, meaning it should be considered that the quantification of brain data may interfere with how users see the world, potentially being less immersed in activities and more outcome driven. Existential authenticity should also be a consideration, where one's self-identity might be affected. The sixth dimension states that consumer neurotechnologies must be distributed fairly, without creating inequalities. Finally, in the absence of proper regulation, a working group of stakeholders should appraise the risks and benefits of neurotechnologies before they become available to consumers. If these dimensions are considered, Kreitmair [139] argues that consumer neurotechnologies would be able to meet their intended purposes of improving lives and experiences instead of having unintended consequences through unconsidered ethical implications.

Additionally, the UN's International Bioethics Committee have recently released (August 2021) a draft report on the ethical issues of neurotechnology [216]. As well as medically issued neurotechnology, the report considered consumer neurotechnologies. The authors outline that whilst there is potential for tremendous benefits, neurotechnologies also hold the potential to damage individuals' privacy, deepen social inequalities and provide tools for the manipulation of individuals. They note that there are few regulations outside of those on medical devices used in research or the medical field, and recommend the introduction of 'neuro-rights' into law. This regards the rights of individuals to retain their integrity, mental privacy, freedom of thought and free will, the right to benefit from scientific progress and freedom of choice on matters related to the use of neurotechnology without any discrimination, coercion and violence.

Along a similar note, Ienca and Andorno [124] discuss how the development of consumer neurotechnology requires the emergence of new human rights, or at least the expansion of already established rights in order to address the emerging challenges of neurotechnology development. Indeed, Ienca et al. [125] included human rights as one of four identified areas that require proactive governance to ensure safe and responsible use of brain data outside of a medical domain, stating that brain data protection needs to be embedded into human rights in order to be included in the international normative framework. Binding regulation, where brain data is given its own category for mandatory data protection was another identified area for regulation. The third identified area was ethical guidelines and soft law, which regards how the collection and processing of brain data is governed. Finally, responsible innovation was the fourth area, which relates to the responsible collection and processing of data (such as validating the technology).

2.5 Tying Together Mental Workload, Personal Cognitive Informatics, and pBCIs

Taking the research outlined so far, we believe that the fields of neuroergonomics, personal informatics, and BCIs fit together seamlessly for a multidisciplinary project that has the potential for a real-world impact. This section will summarise the take-away aspects of each field that underlie my research project.

Firstly, currently available pBCI devices tend to consider cognitive 'workouts' in terms of tracking and understanding brain activity over a short period of time, such as a meditation session, in comparison to a longitudinal perspective. This is perhaps comparable physical activity tracking, which can be tracked for the duration of a specific workout, or more longitudinally over time (such as heart rate data or calories burned over a day/month/year). Whilst workout tracking might be useful, especially as current pBCI devices aim to monitor cognitive activity then provide feedback in order to improve the cognitive state of the user, healthy lifestyles are also contributed to through tracking longitudinal data, such as how many steps we walk each day.

However, in terms of neurotechnology, research has shown that people are particularly interested in collecting and understanding their own cognitive data compared to other data that can be collected through wearable technology [111]. In this respect, there is a lack of research into what type of data might be useful to track longitudinally, how it might be useful, and how users could effectively interact with it.

Therefore, it seems important to develop an understanding of how *pBCI* devices can harness cognitive activity as a form of *personal informatics.*

Further to this, as mental workload is such a meaningful concept in neuroergonomics and pBCI research in order to improve performance at work [16; 240], alongside our argument that mental workload is a meaningful concept in our daily lives and should be considered holistically, it appears to fit seamlessly as a potential candidate for the cognitive activity that may be of use to future pBCI neurotechnology users as a form of personal cognitive informatics.

Chapter 3

Measuring Mental Workload Variations in Office Work Tasks using fNIRS

This study was published in the International Journal of Human-Computer Studies:

Midha, S., Maior, H. A., Wilson, M. L., & Sharples, S. (2021). Measuring mental workload variations in office work tasks using fNIRS. International Journal of Human-Computer Studies, 147, 102580.

3.1 Introduction

This chapter aims to address research question (a):

Can we physiologically track mental workload levels in general work tasks, and what are the practical concerns of doing that?

Our interest is in measuring mental workload longitudinally in our day-to-day lives, but there is a lack of research that explores the physiological measurement of mental workload for tasks that are more relevant to daily life. As mentioned previously, studies that measure mental workload have predominantly been studied in laboratory environments or safety-critical task contexts. In these terms, the study presented in this chapter begins to bridge the gap between tightly controlled laboratory or safety-critical environments and more uncontrolled general work tasks. Hence, this study used physiological methods to measure variations in mental workload levels during naturalistic general office-like reading and writing tasks, which included features commonly associated with office-work environments, namely verbal interruptions, personalised tasks, and coffee consumption. This research could be essential for our understanding of whether it is possible to interpret cognitive activity as people go about their daily lives, and the practical issues that may need to be considered when machine learning research aims to accurately classify mental workload levels for future pBCI technology.

In a working environment, being interrupted whilst undertaking a task is often inevitable. Research has suggested that people at work are interrupted four times per hour on average, and the most common form of interruption is verbally face to face [171], though online distractions are ever-more common [156]. Most research into interruptions has focused on their impact on task performance or completion [155], or into delivering interruptions at a timely stage of a task [126; 22]. The effect of interruptions on mental workload levels, and whether physiological measures can be sensitive to these potential mental workload changes, have not yet been investigated. Yet these factors are intuitively critical for deepening our understanding of the factors involved in tracking mental workload in the real-world.

As mentioned previously in terms of physiological measures, fNIRS is a brain imaging method that uses near-infrared light to infer cognitive states. Because of its ease of use and ability to produce interpretable data in real-world environments [115; 189], fNIRS is arguably becoming reputed as the most promising objective measure for real-world mental workload measurements. Because of this, and the rapid progress with which fNIRS developments are being made, this study uses an fNIRS approach as the physiological measure for investigating the measurement and implications of measuring mental workload across general work tasks. The study is not completely uncontrolled, but instead a stepping stone between tightly controlled studies and more naturalistic environments. By maintaining control over some variables, we should be more confident in the conclusions drawn from the study for which future research can continue to build upon.

Based on the argument outlined above, this study was a naturalistic laboratory study that used fNIRS to measure mental workload variations for personalised reading and writing tasks at different levels of difficulty. Interruptions were incorporated verbally, and coffee drinking was uncontrolled during task conditions.

The following findings were hypothesised:

H1a fNIRS will detect differences in brain activity between conditions that correspond to different mental workload levels of a reading task.

H1b fNIRS will detect differences in brain activity between conditions that correspond to different mental workload levels of a writing task.

H2 fNIRS will detect changes in brain activity corresponding to interruptions.

Through this work we make the following study specific contributions:

- We show that fNIRS measurements can differentiate between reading task levels but saw no significant differences between writing levels (despite self-reported differences).
- We consider fNIRS measurements in terms of spare capacity models to reflect on how interruptions are handled.

These findings are used to provide a higher level contribution:

- Provided a deepened understanding of the factors that will be involved in measuring mental workload in real-world environments.

3.2 Method

3.2.1 Participants

20 healthy participants took part in the study (8 females and 12 males, aged 31 $\pm$ 9.57). Opportunity sampling was used to recruit participants and each participant provided written and informed consent. Participants were eligible for participation if their work included typical office-like tasks, such as reading professional documents. The experiment was approved by the School's ethics board (approval ID: CS-2017-R13) and participants were provided with a £10 Amazon voucher as an inconvenience allowance.

3.2.2 Design

The study had a repeated measures design. There was a reading and writing task, each with three conditions of an 'easy,' 'medium,' and 'hard' difficulty designed to require corresponding levels of mental workload. The easy and medium conditions involved personalised materials, and the hard conditions were a continuation of the medium conditions with an addition of a secondary task designed to overload participants' mental workload according to the Multiple Resource Model.

3.2.3 Materials

Reading task.

The reading materials for each participant were selected by the researcher, personalised to each participant's area of research, work, or study; this meant that participants were presented with different materials, which was aimed at more closely reflecting real-world work environments. The easy condition for the reading task involved participants reading basic material related to their area of research, work, or study. For example, if a participant's PhD project involved studying schizophrenia, the easy condition may involve reading a basic NHS article about the condition. The medium condition involved reading a previously-unread academic journal article, also relevant to participants' individual areas. Using the previous example, the participant may be presented with a scientific journal article about how the brain in affected in people suffering from schizophrenia. For the hard condition, participants continued reading the materials from the medium condition whilst completing a secondary task that competed for the same cognitive resources according to the Multiple Resource Model — this involved counting the amount of times the word 'the' was read. This secondary task was the same for all participants and aimed to represent tasks in which one is searching for words or information whilst reading.

The difficulty of the reading materials was formally assessed using the Flesch-Kincaid Grade Level and Flesch Reading Ease scores [90]. These measures are based upon word length and sentence length and can assess how difficult a piece of text is to read. The easy condition materials were at least 2 levels below the medium condition materials material in both the Flesch-Kincaid Grade Level and the Flesch Reading Ease measures to ensure a definite difference in task demands. All reading materials were presented in an identical format with images removed to reduce the effect of confounding variables.

Writing Task

All writing conditions were conducted in an email format which was addressed to the experimenter, and the difficulty of the conditions was based upon the assumption that an increased amount of required cognitive processes positively correlates with task demand. For the easy condition participants were asked to *"Describe the tasks that you have done so far in this experiment in some detail"*. This was designed to require retrospectively recent memory.

The premise for both the medium and hard conditions were to *"Pretend I have emailed you asking about your area of research. I'm interested in what you research, how you research it and why you research this area. Please reply to that email. You can assume I have a basic but limited knowledge of your field so you will need to explain certain terms to me. I also mentioned that I would be interested in meeting with you to discuss your research."* Wording was altered slightly when required to be relevant to participants.

The medium condition required participants to start by outlining some real days and times that they were available to meet this week and then talk about their research. This required retrospective memory, short term prospective memory and working memory to remember the vast amount of information that was provided. The hard condition required participants to continue with the task as well as outlining some real days and times they would be available to meet the next week. The secondary task for the hard condition involved participants saying 'blah' repeatedly out loud whilst writing. This condition was designed to require retrospective memory, longer term prospective memory and working memory, whilst completing a difficult secondary task that competed for the same resources according to the Multiple Resource Model. This secondary task was the same for all participants and aimed to represent the notion of speaking whilst working.

Interruptions

Verbal interruptions involved the experimenter briefly disrupting the condition with generic conversation, and were added to 3 out of the 6 conditions, counter-balanced between participants. Interruptions lasted for a minimum of 10 seconds and a maximum of 20 seconds.

To further increase ecological validity and make the study environment less controlled, all participants were provided with a drink (coffee, tea or water - if a drink was declined, water was provided on the desk). Drink consumption was permitted as desired except during the baseline conditions.

Baseline Conditions

A fixation cross presented for 1 minute on a monitor was used as a baseline condition at the beginning of both the reading and writing task. Before each further task condition, the fixation cross was used again for 1 minute to allow brain activity to return to baseline.

Mental workload questionnaire

A NASA-TLX workload questionnaire [109] was used to collect subjective mental workload information. The self-assessed questionnaire comprises of six 21 point scales where a rating of 0 equates to 'Very Low' and 20 to 'Very High'. The scales include mental demand, physical demand, temporal demand, performance, effort and frustration.

fNIRS

A wireless fNIRS device (Octamon, Artinis Medical Systems) with 8 channels with a source-detector distance of 3.5cm measured oxygenated (O2Hb) and deoxygenated (HHb) haemoglobin across the PFC (Figure 3.1).

The wavelengths used were 760 and 850nm with a differential pathlength factor of 6 and a sampling rate of 10Hz.

3.2.4 Setup and Procedure

A standard office set-up was created. Participants sat behind an office desk and in front of a computer monitor and keyboard. A non-transparent board was placed between the participant and experimenter to provide a sense of open-plan office form of semi-privacy. In order to identify the times at which the verbal interruptions occurred, a GoPro Hero4 Silver placed in an audible protective case was used. This was placed inconspicuously behind participants and recorded their frame and monitor.

Study Procedure

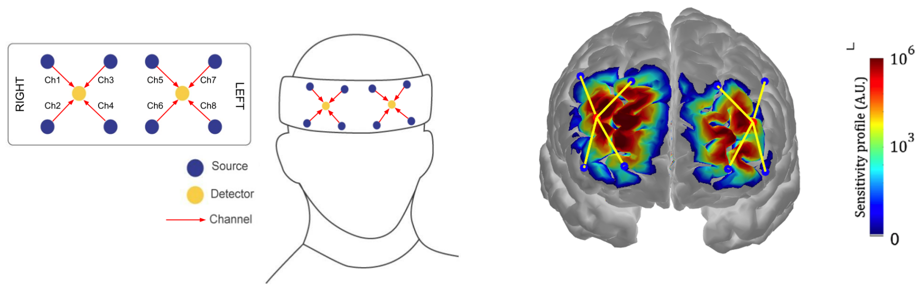
Participants were first provided with a drink and seated at the desk where the study instructions were given and informed consent was provided. The fNIRS device was fitted and the GoPro was started.

The reading task was completed first due to the writing task being partially based on the reading task. When the study started, participants first stared at the fixation cross for 1 minute before the easy reading condition began. The condition lasted for 5 minutes, and when time was up participants immediately filled out a NASA-TLX questionnaire. After the questionnaire was completed, the fixation cross appeared again for 1 minute before the next reading condition began. The order of the medium and hard reading conditions were counterbalanced across participants and all conditions lasted for 5 minutes. Once the second condition was completed, the NASA-TLX questionnaire was administered immediately again, followed by the 1 minute fixation cross and the final reading condition which was again followed by the NASA-TLX questionnaire. After the reading conditions, the writing conditions started and followed the same format as the reading conditions.

3.2.5 Data Analysis

NASA-TLX

Friedman tests were run to investigate whether there was a significant difference in ratings between the easy, medium and hard reading and writing tasks; post hoc analysis was conducted with Wilcoxon signed-ranks with a Bonferroni correction applied.



(a) Sensor layout for the Octamon fNIRS device with a source-detector distance of 3.5cm.

(b) Visual representation of the sensitivity profile of the channels. Created using AtlasViewer [2].

Figure 3.1: Sensor placement and sensitivity

fNIRS Measurements

Raw data was exported to Homer2 fNIRS processing package [121]. Data was converted into changes in optical density, and motion artifacts were corrected using a Wavelet filter (iqr=1.5) and a bandpass filter (0.5 LPF and 0.01 HPF). Physiological noise was reduced using a Principal Component Analysis (nSV=0.97) and concentration changes in O2Hb and HHb were calculated using the Modified Beer-Lambert Law.

Baseline correction was performed by subtracting baseline mean values from the task data; the first baseline condition from both the reading and writing tasks were used and subtracted from their respective tasks. To test whether there were significant differences in brain activity between the easy, medium and hard difficulty reading and writing tasks, one-way repeated measures ANOVAs were conducted using the average values from the first 2.5 minutes of each condition before any interruption occurred.

For the interruption analysis, the interruption timings were marked down from the video footage and added as stimuli in Homer2. Data was baseline corrected in the same way as the task data. To compare brain activity during the task compared to during the interruptions, paired t-tests were used to compare 10 seconds of interruption data against the previous 10 seconds of task data for the participants that were interrupted in the same conditions as the interruptions. All interruption stimuli were shifted by 2 seconds after the onset of the interruptions to account somewhat for the temporal delay of fNIRS measurements.

When a region of the brain becomes activated, cerebral blood flow increases to meet the increase in oxygen demand; this is known as the hemodynamic response and is reflected by an increase in O2Hb and a decrease in HHb [201]. Measurements of O2Hb alone are vulnerable to physiological noise which risks false conclusions about neural activity being drawn [186; 116]. Whilst measurements of HHb are less affected by these confounds (page 364) and most highly correlated with other brain imaging methods [122] the strongest indicator of functional brain activity is when there is an increase in O2Hb corresponding with a decrease in HHb [116], and thus this is what will be our main focus in the results.

Data from 2 male participants were excluded due to technical difficulties making the final analysis include data from 18 participants. All fNIRS measurements are reported in micromoles (μM). Post-hoc analysis for the fNIRS data was conducted using a Bonferroni correction.

Coffee Consumption

The intention was to analyse the effect of drink consumption on mental workload levels. However, this analysis was not possible for this study for two reasons. Firstly, the period of time that it took for participants to take a drink was very short (a couple of seconds), meaning that because of the temporal delay of fNIRS, we could not be certain that we would capture the drinks data in the analysis, reducing the validity of the results. Secondly, there was such large variability between participants in the frequency of drink consumption; for the participants who drank very frequently, it would have been difficult to extract the drinks data from the task data, and left the task data very fragmented. Therefore, the drink consumption data was not analysed, and instead left in as a factor that reflected real-world working conditions.

3.3 Results

3.3.1 NASA-TLX ratings

Figures 3.2 and 3.3 display mean subjective scores for the Mental Demand, Effort and Performance subscales for reading and writing respectively.

A Friedman test revealed a significant effect of condition on mental demand ratings for the reading task, $\chi^2(2)=18.250$, $p=0.001$. Post hoc analysis with Wilcoxon tests was conducted with a Bonferroni correction applied which resulted in a $p=0.017$ significance value for all NASA-TLX post hoc analyses. This showed that the easy and medium reading conditions were rated significantly lower than the hard reading condition ($Z=-3.335$, $p=0.001$ and $Z=-2.519$, $p=0.012$ respectively) but did not significantly differ to each other. Mental demand ratings also showed significance in the writing task ($\chi^2(2)=14.464$, $p=0.001$), where the easy and medium conditions were both rated lower than the hard condition ($Z=-3.463$, $p=0.001$ and $Z=-2.872$, $p=0.004$ respectively) but did not significantly differ to each other.

There was a significant effect of condition on physical demand ratings for the reading task, $\chi^2(2)=17.211$, $p=0.001$. Wilcoxon tests showed that the easy and medium reading conditions were rated significantly lower than the

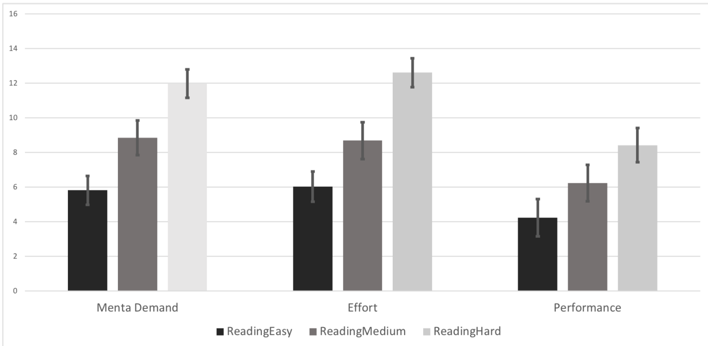


Figure 3.2: Average NASA-TLX ratings across the reading task conditions for Mental Demand, Effort and Performance sub scales. All error bars represent standard error of the mean.

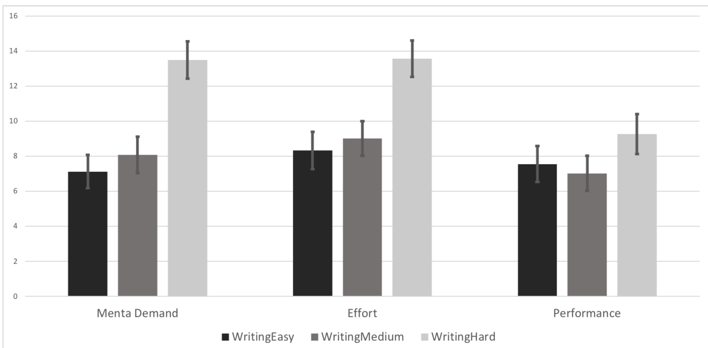


Figure 3.3: Average NASA-TLX ratings across the writing task conditions for Mental Demand, Effort and Performance sub scales.

hard condition (Z=-2.958, p=0.003 and Z=-2.534, p=0.011 respectively) but were not rated significantly different to each other. Physical demand was not significant for the writing task. There was also no significant effect of condition on temporal demand or performance ratings for the reading or writing tasks.

A significant effect of condition on effort ratings for the reading task was found, $\chi^2(2)$=25.054, p=0.001. Post hoc analysis revealed that the easy and medium conditions were rated as significantly lower than the hard conditions (Z=-3.830, p=0.001 and Z=-3.604, p=0.001 respectively), but again did not significantly differ to each other. The Friedman test was also significant for the writing task, $\chi^2(2)$=22.243, p=0.001. Wilcoxon tests revealed that the easy and medium conditions were rated lower than the hard difficulty condition (Z=-3.428, p=0.001 and Z=-3.732, p=0.001 respectively) but did not differ to each other.

A significant effect of condition on frustration ratings for the reading task was revealed, $\chi^2(2)$=18.613, p=0.001. The post hoc analysis showed that the easy and medium conditions were rated significantly lower than the hard condition (Z=-3.416, p=0.001 and Z=-3.316, p=0.001 respectively), but not significantly different to each other. Effort was also significant for the writing task, $\chi^2(2)$=15.254 , p=0.001, where the easy and medium conditions were rated significantly lower than the hard condition (Z=-3.157, p=0.002 and Z=-3.465, p=0.001 respectively) but again were not significantly different to each other.

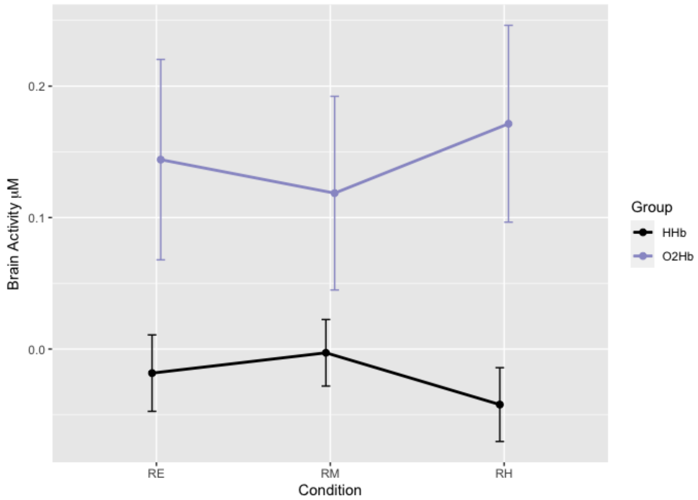


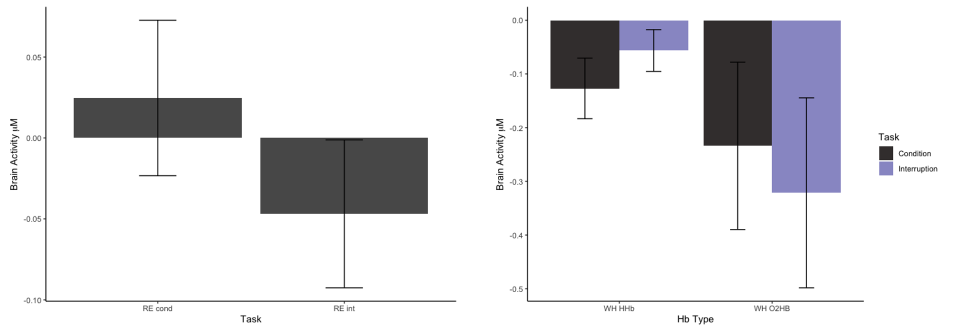
Figure 3.4: fNIRS results for the reading task show a mirroring pattern of increased O2Hb and decreased HHb during the hard difficulty condition - Channels 5-9.

3.3.2 fNIRS data

Condition analysis On an individual channel basis, only channel 7 HHb (see Figure 3.1) showed a significant ANOVA result ($F(2, 34)=4.258$, $p=0.022$), where means showed that reading medium had the lowest brain activity, followed by reading easy, and reading hard had the highest brain activity, though this did not reach significance in the Bonferroni post hoc test.

When averaging across the different sides of the PFC, the ANOVA showed a significant reading result for channels 5-8 (left side of the PFC) for both O2HB ($F(2, 34)=3.400$, $p=0.045$) and HHb ($F(2, 34)=3.425$, $p=0.044$). For both O2Hb and HHb, the means showed that like channel 7, reading medium had the lowest brain activity followed by reading easy, and reading hard had the highest brain activity (Figure 3.4), though Bonferroni correction did not reach significance.

No writing results reached significance in the condition analysis.



(a) fNIRS results show a drop in HHb during the interruption of reading easy compared to task alone - Channel 7.

(b) fNIRS results show decreased O2Hb and increased HHb during interruptions of the hard difficulty writing task - Channel 1-4.

Figure 3.5: fNIRS measurements during task completion compared to during the interruption periods.

Interruption analysis On an individual channel level, the paired t-tests showed that for channel 6, those who were interrupted during the reading easy task (n=8) had significantly higher O2Hb brain activity during the task compared to during the interruption ($t(7)=3.119$, $p=0.017$). On the contrary, channel 7 showed significantly more HHb (indicating less brain activation) during the reading easy task compared to during the interruption ($t(7)=2.525$, $p=0.040$) (Figure 3.5a). Channel 7 also showed significantly less HHb levels (more brain activity) for participants (n=8) during the writing hard task compared to during the writing hard interruptions ($t(7)=2.749$, $p=0.029$).

When considering the averages for the sides of the PFC, channels 1-4 (right side of the PFC) for both O2Hb and HHb were significant for writing hard ($t(7)=2.496$, $p=0.041$ and $t(7)=2.514$, $p=0.040$ respectively). Both the O2Hb and HHb showed, like channel 7, more brain activity during the writing hard task compared to during the interruption during the writing hard task (Figure 3.5b).

3.4 Discussion

The study aimed to investigate the measurement of mental workload using fNIRS in tasks and environments more relevant to daily life. The study was a controlled laboratory study, but was designed to reflect several aspects of real-world working environments, including relevant tasks, verbal interruptions, and unconstrained coffee consumption. The findings can provide guidance for progressing research into more general environments, and may be especially useful for machine learning research as it progresses towards measuring mental workload outside of laboratory environments [27].

3.4.1 The Sensitivity of fNIRS

Results showed support for the ability of fNIRS to detect mental workload levels in reading tasks (H1a) but not writing tasks (H1b). For the reading task, fNIRS could detect significant differences between conditions in the left side of the PFC and the results aligned with the subjective ratings of mental workload in terms of the hard condition showing the highest mental workload. Subjective ratings of mental workload showed no significant differences between the easy and medium conditions, and fNIRS measurements showed only small differences between these conditions compared to the hard condition, also supporting the sensitivity of fNIRS (H1a).

The reading results are supported by the findings from other studies. Because of the different sizes of participant's heads and the fixed optode locations on the fNIRS device, we can broadly assume which channels correspond to the different areas of the PFC via the 10-20 electrode system from the EEG field [127; 116]. Reading comprehension is heavily associated with the PFC (as well as temporal regions), namely Broca's area located in the left inferior frontal gyrus, and thus the consensus is for left hemispheric dominance in reading comprehension [175; 25; 193]. Our results were significant for channel 7, which is in the left hemisphere and where Broca's area is expected to be located [220], in line with the areas of activation one would expect to find. The results were also significant for channels 5-8, which is the left hemisphere, further corresponding to previous literature, including fNIRS studies into reading comprehension [69]

As there was strong evidence for fNIRS detecting differences in mental workload levels between conditions for channels 5-8 with significant results for both O2Hb and HHb, it seems sensible to infer that averaging across channels captured the activation area across participants, which may not have been captured fully by channel 7 if it didn't exactly correspond to the inferior frontal gyrus in all participants.

Regarding the non-significant post hoc tests of the ANOVA, Bonferroni is a conservative test as it attempts to control the overall alpha level. As ANOVA results for both O2Hb and HHb were significant, we consider this as strong supporting evidence for fNIRS detecting differences [188] between reading difficulties, and hence our focus is on this global effect.

As means for brain activity consistently showed the medium condition corresponding to slightly less activity than the easy condition, and non-significant differences in subjective ratings between these conditions, this is an interesting finding for the challenges associated with objectively tracking of mental workload in the workplace (and beyond). Even if task demands were harder for the medium condition, it appears that participant feelings could have impacted on the results. As the medium condition had more stimulating materials, it is expected that participants felt more engaged with the task. Indeed, previous work found that mental workload ratings were lower for demanding tasks when participants were more engaged [117]. Similarly, Lukanov et al. found that, for a insurance claim form, participants preferred the user interface condition that objectively and subjectively generated the highest levels of mental workload [151], and thus emotional factors could be a challenge for objectively measuring mental workload in the workplace.

In contrast to reading, fNIRS did not detect significant differences between conditions in the writing task, not aligning with the subjective results and not supporting H1b.

Writing organisation in the brain appears to represent a complex human function that involves several language sub-components, and thus localisation in the brain is highly individualised between people [150]. Nevertheless, writing localisation is thought to heavily involve the frontal lobe (and the anterior parietal lobe), more specifically the posterior part of the middle and superior frontal gyri (Exner's area) [150; 134]. The fNIRS device in the current study measured isolated activation in the PFC, meaning activation from the 'writing centre' was not covered. This finding, where different levels of mental workload for the writing task were not distinguished in the PFC, does not only hold relevance for the current study, but also challenges perceptions from the wider HCI field.

FNIRS studies of mental workload most often measure cognitive activity from the PFC [185; 153; 152; 209; 208; 242]; there is a consensus that mental workload will consistently be exhibited and measurable in the PFC. However, the results from the writing task suggest that the processing involved when task demand increased and subsequent mental workload increased between the hard and the easy and medium conditions (as found in the subjective ratings) may not have been measurable from the PFC.

As the writing task required a complex amount of neural processing and cognitive processes, it seems that it was not possible to capture the full picture of mental workload level from the PFC alone. Working memory cognitive load is measurable from the PFC [89; 215], but the writing conditions might not have differed significantly from each other in the working memory aspect of the writing task like we intended with the study design, and the combination of cognitive processes that increased mental demand for the hard condition might have been localised outside of the PFC. This is supported by some of our recent work [11] which did not find differences in oxygenation in the PFC for a visual search task despite significantly different subjective mental workload ratings; here, mental workload might have been detectable in the occipital and parietal lobes [215], and did not involve the PFC enough to detect physiological changes in mental workload level. Thus, future studies of realistic or real-world tasks should consider carefully which brain areas mental workload might be most represented in, and perhaps measure multiple lobes to gather a richer insight into brain activation as these types of task come with more cognitive complexities compared to laboratory studies [14]. Future studies could also benefit from further investigating varying levels of mental workload and specific cognitive task demands for tasks and activities relevant to daily life. This perhaps would also explain why papers using machine learning to classify mental workload levels [199; 114; 44] typically achieve fairly low levels of accuracy. In support, it has been shown that measuring a larger neural area resulted in a higher accuracy of mental workload classification [198]. That being said, the PFC may more often than not provide a reasonable insight into mental workload levels due to its involvement in a wide variety of tasks. It is notable that the reading results support the finding that changes in mental workload for reading are detectable in the PFC, where more complex reading tasks are associated with increased neural activity

[131; 130].

A limitation of the study could be that the data was analysed in 2.5-minute blocks which is likely to contain more artefacts compared to shorter trials [246] that are often seen in laboratory studies. A main research area in pBCIs and neuroergonomics is to use fNIRS to measure mental workload of workers in safety-critical jobs and to develop aids to improve the safety of these jobs based on their mental state. Our aim is related to this, in the sense that we wish to investigate mental workload levels in office workers and this data might progress to aid improvements to working habits and lives. These types of research aims (in addition to many other research areas using fNIRS) require long-time continuous monitoring of brain activity, which is an acknowledged advantage of fNIRS [189]. With rigorous processing of the data, the noise should not impact the validity of the data, and longer blocks may reveal a more representative picture of brain activity compared to a short snapshot. We did opt to shorten the analysis from 5 minutes of data to 2.5 minutes of data with the aim of somewhat ensuring control over the data considering the study was still essentially lab based.

3.4.2 Interruptions

In terms of the effect of verbal interruptions on fNIRS measurements (H2), there was less brain activity in the right side of the PFC during the interruption of the writing hard condition, compared to before the interruption, and this was significant for both O2Hb and HHb. If we consider this in relation to spare capacity [204], as the writing hard condition was subjectively rated as requiring the most mental workload, there might not have been enough cognitive resources available to take on the interruption concurrently with the task. This means the interruption would have become a primary task which was less demanding than the writing hard condition. Whereas for the reading easy condition, HHb levels (which are interpreted a bit more cautiously alone) showed an increase in brain activity during the interruption compared to during the task. In relation to the model, this could be explained by the notion that there was spare capacity during the reading easy condition which meant that responding to the interruption could have been achieved through multi-tasking which increased mental demand and hence mental workload. Observationally, the authors note that verbal responses to interruptions during the reading easy conditions tended to be briefer and more distractable, whereas in the writing hard participants could not keep saying 'blah' and respond to the interruption so they would pause the task to give a proper response.

Even though results were only significant for two of our conditions, perhaps due to the small number of participants for each interruption, they do contribute to a deeper understanding of the factors involved in physiologically measuring mental workload in the workplace which are not encountered during controlled lab studies. The results emphasise that changes in mental workload levels are not 'black and white'; instead they often depend on situational factors which might be different from one person to the next [203]. This also means that it is difficult to evaluate the effectiveness of physiological measures because determining changes in mental workload in these situations depends on subjective interpretation. Physiological measures might not reflect the interpretation made about a person's mental workload but that does not necessarily mean that the measurement is wrong. Such variability in factors contributing to mental workload levels and the different responses between people could mean that future studies might benefit from considering results on a participant by participant basis. Additionally, whilst we have incorporated natural interruptions into the study design to gain an understanding of challenges to do with subtle within task variations of mental workload whilst completing a single task, workplaces are increasingly made up of multi-tasking activities [154] which most likely means that physiologically measuring mental workload in the workplace will come with increased complexity. It should be noted that the sample size was limited for the interruption analysis; with more data to analyse the statistical power and validity would have been increased. We do believe our results have opened an interesting area for discussion and further research area on the effect of interruptions on mental workload levels and how these measurements can be dealt with in real-world settings.

We further increased ecological validity by including uncontrolled drink consumption. Because this data was messy, due to participants drinking at different frequencies and during different conditions, it was not possible to analyse the effect of drink consumption on brain activity. This type of interruption data, however, could prove to be valuable if incorporated plausibly in future studies, as it could provide insight into whether different types of interruptions seem to follow the same trend in which data can be understood in relation to spare capacity models, or whether different factors need to be considered.

Finally, continuing to bridge the gap between controlled lab studies and real-world studies is hugely important. Ladouce suggested that true cognition and its complexities can often only be understood correctly when examined in ecologically valid environments [141], and mental workload appears to belong to this category. Progressing this research to in-the-wild studies of mental workload will enable further understanding of the research and challenges associated with the sustained measurement of mental workload in daily life with fNIRS as a potential candidate. Our further work aims to make progress towards being able to track cognitive activity as a form of personal informatics.

3.4.3 Conclusion

Through using personalised tasks and verbal interruptions in a workplace-like setting, we show that fNIRS placed over the PFC alone was able to detect the differences in mental workload experienced by participants during personalised reading tasks, but was not sensitive to the reported differences in writing tasks. This finding could be due to the PFC not exhibiting mental workload levels for all tasks. Thus, careful consideration of optode placement over a larger region of interest is emphasised for naturalistic studies, and this highlights challenges surrounding the sensitivity of physiological measures in real-world environments. Verbal interruptions appeared to cause within task mental workload variation, causing increased load if done in parallel with tasks and decreased load if becoming the primary task temporarily. These findings demonstrated the complexity of mental workload as concept that is non-quantifiable and often affected by situational factors reliant on interpretation. With the goal in mind of the sustained objective monitoring of mental workload in daily life for self-improvements, further work is needed to establish the sensitivity of fNIRS for further general tasks and to build on understanding of the factors involved in these measurements for future machine learning studies.

Future work from this study generally implicates the field of machine learning. A different but related direction for this research concerns the field of personal informatics, in terms of investigating the *use of* and *interaction with* the data once machine learning has the ability to accurately measure mental workload levels from physiological sensors. In this respect, understanding mental workload as a form of personal informatics will be the focus of studies 2 and 3 in this thesis.

Chapter 4

A Longitudinal and Holistic Approach to Mental Workload

4.1 Introduction

Study 2 aimed to address research questions (b), (c), and (d):

(b) *Can a longitudinal and holistic approach to mental workload improve understanding of how mental workload could be valuable as a form of personal informatics, and mental workload as a concept itself?*

(c) *How can objective mental workload tracking data be meaningfully communicated to users?*

(d) *What should be ethically considered when developing mental workload pBCI devices, or neurotechnology in general?*

Stemming from the conceptualisation of mental workload in the literature as a notion only relevant for isolated work tasks in predominantly safety-critical environments, there is a lack of a more holistic understanding regarding the function of mental workload in daily life, and hence what types of goals we could be setting in that regard.

As mentioned previously, this research question is multifaceted. This is because in order to understand mental workload holistically, specifically in relation to the development of pBCI devices, there is a need to explore individual perspectives about the concept as well as developing a more longitudinal view.

An empirical study was therefore conducted that explored a holistic approach to mental workload from a quantitative and a qualitative perspective. The same participants took part in both phases. In the quantitative phase, participants were subject to a 5-day longitudinal data tracking, in which subjective mental workload ratings and diary data were inputted into an app at frequent intervals from Monday-Friday. Alongside this, mobile phone usage data was continuously collected throughout the week, and evening questionnaires were completed each day, which regarded aspects such as perceived stress and mood levels. The aim of the questionnaires was to quantitatively investigate how mental workload impacts certain aspects of our lives.

There is limited research on the consequences that mental workload has on aspects of our daily life. Literature also often refers to physical workload and mental workload interchangeably, meaning it is often unclear if research is referring to the sheer amount of work that an individual is required to complete (workload) or whether the subjective experience of responding to the task demands are considered as well (mental workload). Differentiating between these terms is essential in the human factors research field [240], but seems to be less practiced in other fields.

However, there is some evidence that higher mental workload levels are associated with feelings of fatigue, shown by rail workers feeling more fatigued immediately after their work shift [70]. Additionally, a large laboratory study run over a five night period subjected participants to three daytime periods of either moderate mental workload levels or high mental workload levels before monitoring nighttime sleep [101]. The authors found that higher mental workload levels during the day was associated with increased feelings of fatigue at the end of the day and longer sleep onset latencies. In another study on midwives at work, some factors of mental workload were found to predict mood, such that higher task demands, effort, and frustration contributed to more negative moods during their work shift [181]. To the best of our knowledge, the relationship specifically between mental workload and food cravings has not been researched, but some research has used high mental workload situations to indicate feelings of stress and then investigated how stress is related to food cravings and intake [99]. In this sense, links between stress and food cravings have previously been identified [190]. Finally, it has been established that there is a relationship between mental workload and stress [8]. This is in terms of mental workload and stress models having overlapping aspects that mean physiological measurements often struggle to differentiate between them, especially as high mental workload tasks can often induce feelings of stress [8]; the longer term impacts of mental workload on stress levels, however, do not appear to be established.

Thus, in order to understand how pBCI devices that track mental workload could facilitate life improvements, it is important to develop a deeper understanding of the impact of mental workload in daily life.

The qualitative phase took place the week following the quantitative phase and involved in depth interviews about participants' perceptions of mental workload, including personal experiences, conceptualisation, contextualisation, and ethical considerations. The same participants were included in the qualitative research as the quantitative tracking as it was speculated that including participants that had specifically considered their own mental workload data in their lives would enable a deeper and richer insight during the interviews.

Study specific contributions from the qualitative phase included:

- We identify mixed and changing perceptions of mental workload.
- We identify the importance of fluctuating between mental workload levels in daily life in terms of performances, perceptions, and wellbeing, and reasons why this might not always be possible.
- We identify the characteristics of metaphors that resonate with potential neurotechnology users, as well as colours, shapes, and descriptors.
- We identify ethical concerns and perceptions of potential neurotechnology users regarding privacy, data validity, personal identity, data misinterpretation, and enforced tracking.

Study specific contributions from the quantitative phase included:

- We develop a broader view of the mental workload levels and transitions experienced in daily life and how these can differ from qualitative preferences.
- We further develop an understanding of the daily life impact that mental work has, in terms of distractions and perceptions related to wellbeing.

These findings are used to provide higher level contributions:

- Developed of the Mental Workload Cycle as a model of mental workload. The Cycle both indicates how mental workload data can be useful as personal informatics, and develops understanding of mental workload as a concept. (Study 3 and Study 2)
- Produced design recommendations for the development of pBCI mental workload devices. (Study 3)
- Expanded ethical considerations specifically for the development of pBCI devices. (Study 3)

The data from the qualitative phase was analysed first, and the insights from that study were used to guide the analysis for the quantitative phase. More specifically, the qualitative personal experiences section identified the Mental Workload Cycle. During the interviews, participants frequently reflected on and referred back to their experiences during their time in the quantitative phase. It was therefore decided that the analysis for the quantitative phase would be an opportunity to quantitatively investigate the qualitative experiences, as well as further develop the model of the Mental Workload Cycle. Thus, the results relating to the Cycle are presented in the same chapter (Chapter 5); the qualitative findings relating to the design and ethical perceptions surrounding mental workload pBCI devices are outlined in Chapters 6 and 7 respectively.

4.2 Quantitative and Qualitative Experiment Design

4.2.1 Participants

19 purposive participants took part in the study, recruited by opportunity sampling. Participants responded to advertisements put out through email groups and social media channels. To be included in the studies, participants were required to (a) complete office work as part of their jobs, (b) be Android users, and (c) have no clinical history of anxiety or depression. Out of the 19 participants, 9 were based in academia and 10 were industry workers (see Table 4.1). Ages ranged between 21 and 45. Office workers were chosen as a sample considered representative of our wider focus on tracking cognitive activity in daily life as a form of personal informatics; work tasks in this sample were considered to be primarily cognitively based as opposed to shop or factory style work which may e.g. include more influence from physical workload and fatigue. Ethical approval was granted for the studies [CS-2019-R13] and all participants provided informed consent before data collection began.

Participants were automatically awarded £75 as remuneration for participation in both studies. If their data response was considered good by the researcher, an additional £25 was provided. All participants did receive the full £100 for their participation.

Table 4.1: Table showing participants by ID, along with the occupation, age, and self-identifying gender.

Participant	Occupation	Age	Identify as
P1	PhD Candidate	25	Male
P2	Regional Account Manager – Field Sales	24	Male
P3	PhD Candidate	45	Female
P4	Post-Doctoral Researcher/Teacher	32	Male
P5	Research Fellow	35	Male
P6	PhD Candidate	28	Female
P7	PhD Candidate	30	Male
P8	MSc Candidate	27	Male
P9	PhD Candidate	27	Female
P10	Commercial Finance Manager	44	Female
P11	PhD Candidate	31	Male
P12	Copywriter	33	Male
P13	Ecologist	26	Female
P14	Business Support Administrator	32	Female
P15	Software Engineer	21	Male
P16	Programme Support Officer	41	Male
P17	Software Engineer	33	Male
P18	Voluntary Deputy Services Manager	35	Female
P19	Senior Health Economics Manager	33	Male

4.2.2 Materials

All contact between the researcher and participants was virtual. Participants were provided with a total of 5 documents via email at different stages. Firstly, before committing to participation in the study, potential participants were provided with an information sheet (Appendix A) which contained details about the two phases, such as the different measures that would be taken and the advantages and disadvantages of taking part. A privacy notice (Appendix A) was also provided at this time with information about how participants' data would be protected alongside detailing their rights and their risks. Once participation in the study had been confirmed, written consent was obtained (Appendix A).

After this, participants were provided with a 'pre-study' document (Appendix A). This document provided participants with more insight about the background behind the study, the research aims from the researcher's point of view, and what we hoped to 'get' out of each participant.

The pre-study document aimed to maximise the contribution of each participant by increasing their understanding of the study and encouraging them to consider mental workload in their lives before being probed in depth about their opinions, experiences, and concerns during the qualitative phase. We considered it beneficial to share the information in the pre-study document to increase the richness and validity of data, and ensure that mental workload was considered in a way that aligns with what is established about the concept in the literature.

In terms of the background behind the study, an introduction to mental workload was outlined, in which it was stated that no definition is universally agreed, but the characteristics of mental workload were provided. It mentioned how mental workload is not the same as stress (in which the researchers have previously experienced participant confusion), and briefly described how we are interested in whether mental workload might be useful for us to track in our daily lives, in a similar way to which we can track physical activity.

The research aims of the study were outlined following on from the study background, including how we wished to investigate mental workload variations in daily life, the impact of mental workload, and how it is qualitatively conceptualised.

In regards to what was asked from each participant, it was described that aside from full commitment to the study, we were creating the opportunity to have very in depth insights into mental workload as a concept, and thus we hoped participants would 'tune in' to their mental workload experiences and opinions during the quantitative data collection phase so that they could provide insightful data for the interviews. The different interview sections were then outlined (personal experiences, conceptualisation, contextualisation, and ethical considerations - described below) and examples were provided.

The day before data collection began, participants were provided with a 'practicalities' sheet (Appendix A). This sheet was a help sheet designed to provide practical information about the measures being taken (outlined below) and situations participants may encounter during the quantitative phase. Along with the practicalities sheet, software instructions were also provided which detailed how to set up the measures for the quantitative phase.

Once this information had been provided, data collection began. Information about participants was gathered using a participant bio questionnaire (Appendix A), which asked participants specific details about themselves. This included their age, gender identity, their area of work or study, their work or study responsibilities, and their usual working hours.

The measures outlined above were administered in relation to both phases. Outlined below are the phase-specific measures and procedures.

4.3 The Quantitative Phase

4.3.1 Design and Materials

Mental Workload Ratings App

An app was designed to collect subjective mental workload level data from each participant (Appendix B) . It operated only on Android phones, hence the inclusion criteria that required Android users. Participants activated the app each morning upon waking, and closed the app fully each evening before sleeping. Upon opening the app for the first time, participants were instructed to input their general working hours. This was because the app changed the frequency of notifications depending on work hours; during working hours, the app requested data entry every 30 minutes, and outside of working hours, it requested data entry every 1 hour. Once working hours had been inputted, participants could begin entering their mental workload data. The app stored the working hour information, such that participants did not have to enter the information each time the app was opened. When the app was opened at the beginning of each day, participants selected 'start' in order to start their day of data input. The first question participants were asked was 'What is your current level of mental workload?'. The response options were those of the ISA scale [36] (outlined in the literature review). The display showed a rating number and the corresponding mental workload key: 1-under-utilised, 2-relaxed, 3-comfortable, 4-high, 5-excessive. Participants selected the response that applied to them and were taken to the end page which outlined how long it would be until they received the next notification.

After the first rating of the day, participants were presented with three more questions each time they entered data into the app. This was because these questions related to the time period between the current rating and the last rating, and thus would not have been appropriate to include for the first rating of the day that had no previous ratings to refer to. The second question asked 'What has you overall mental workload level been since the last rating?' and again used the ISA scale. The third question related to performance, and there was a 5-point likert scale for participants to respond to. The question asked 'What would you rate your overall performance since the last rating?' and the likert scale outlined: 1-very poor, 2-poor, 3-average, 4-good, 5-very good.

The final question requested diary data input, and asked participants to 'Please use the space below to report a summary of the tasks performed since the last rating.' Participants could then provide text entry and were advised to use a bullet point entry style in the pre-study document. The notification times started from the time of the last data entry. Participants were only notified once when it was time to provide the next data entry, and notifications were standard in that they popped up on the participant's phone screens. Each data entry could be completed quickly, taking approximately 1 minute.

In the pre-study document, it was outlined that the ISA scale is a widely used scale, and like all available mental workload scales, it had been designed specifically for work tasks. Participants were therefore advised to make logical decisions about how to enter ratings referring to activities outside of work tasks. An example was provided outlining: 'E.g. if you rate your mental workload level as Under-utilised, you should feel that you have very many spare mental resources that are not being allocated to the activity/activities that you are doing/have done.'

Stress Questionnaire

The Perceived Stress Scale (PSS) [55] (Appendix B) was used as a daily dependent measure alongside the other questionnaires outlined below. It is the most widely used measure to assess levels of perceived stress, and is a measure of how stressful situations in a person's life are appraised as stressful [55]. Questions relate to how unpredictable, uncontrollable, and overloaded participants find their lives currently. Questions apply to general situations instead of context-specific, and thus can be applied to any population. The scale has 10 questions and asks about participants' feelings and thoughts over the past month. For the purpose of the current study, this was changed to ask about participants' feelings and thoughts over the past day. For example, 'Today, how often have you been able to control irritations in your life?'. Answers were on a 5-point likert scale: 0-never, 1-almost never, 2-sometimes, 3-fairly often, 4-very often (Appendix B).

Mood Questionnaire

A widely accepted measure of current mood is the Profile of Mood States questionnaire (POMS) [61]. The length of the original scale was 65 items, which was time consuming. A short version was later developed, consisting of 37 items [61]. However, these scales are under copyright, meaning permission and payment are required for use. Additionally, many experts believe these scales have been superceded by other mood measuring scales [202].

An abbreviated Profile of Mood States was developed by Grove and Prapavessis (revised POMS) [104] (Appendix B). The revised POMS questionnaire has been thoroughly validated as a measure of current mood

[214; 104]. It has 40 items and referring to measures of tension, depression, fatigue, vigor, confusion, anger, and esteem-related affect. Total mood disturbance is measured by combining these measures. It is freely available and was the measure of mood used in this study. The questionnaire asks participants to select the answer that best describes how they feel 'right now'. The answer options are on a 5-point likert scale: 0-not at all, 1-a little, 2-moderately, 3-quite a lot, 4-extremely.

Fatigue Questionnaire

The Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F) [144] (Appendix B) was used as a measure of fatigue and energy levels. It consists of 18 items which measure participants' subjective levels of fatigue and energy. It compares favourably to other measures of subjective fatigue [144] and has high internal reliability [144; 202].

The scale again asks participants to circle the response which best describes how they are feeling 'right now'. The responses options use a scale from 1-10, with the extreme answers on each side of the scale. For example, 'not at all sleepy' is placed beside a rating of 1, whilst 'extremely sleepy' is placed beside a rating of 10. Participants circle the number along the scale relating to how they feel in that moment.

Appetite Cravings Questionnaire

Two food craving questionnaires have been widely used and validated, called the Food Cravings Questionnaire-Trait (FCQ-T) and the Food Cravings Questionnaire-State (FCQ-S). [43; 169]. The FCQ-T measures food cravings as a stable and unchanging traits within individuals or populations. The FCQ-S measures food cravings as state-dependent, meaning that cravings are measured in terms of variable responses to specific and changing situations, or psychological and physiological states [169]. For the purpose of this study, the FCQ-S was used (Appendix B).

The FCQ-S has 15 items relating to how participants feel 'at this very moment'. It has a 5-point likert scale: 1-strongly disagree, 2-disagree, 3-neutral, 4-agree, 5-strongly agree. The scale measures 5 factors relating to food cravings, and 3 questions relate to each factor: an intense desire to eat, anticipation of positive reinforcement that may result from eating, anticipation or relief from negative states and feelings as a result of eating, obsessive preoccupation with food or lack of control over eating, and craving as a physiological state.

Sleep Diary Questionnaire

Morgan et al [1] developed a daily sleep diary that is recommended for use by the NHS in order to monitor the quality and quantity of sleep. Sleep diaries keep track of sleep quality over time, whereas sleep questionnaires are usually used in sleep clinics as a one time measure [123]. There are several types of sleep questionnaires and diaries which vary in length and are often used to diagnose sleep disorders.

The sleep diary developed by Morgan et al was used in the current study due to its convenience for completion (Appendix B). The diary has 8 questions that require text input. For example, 'How long did you spend in bed last night?', for which participants enter the amount of time in hours and minutes. The last question on in the diary is, 'How would you rate the quality of your sleep last night?' and the scoring system is a likert scale

[1] https://www.nhs.uk/livewell/insomnia/documents/sleepdiary.pdf

from 1-5, where 1 is outlined as very poor and 5 is outlined as very good. The diary has different columns for the different days of completion.

Post-Study Document

A post-study document (Appendix B) was provided that thanked the participants for their participation in the quantitative phase and provided instructions for sending their data to the researcher and uninstalling the downloaded software. Specifically, participants were asked to email the questionnaires to the researcher. For the mental workload ratings app, participants' data was saved to a .csv file that could be located through the search function and emailed to the researcher. Once receipt of the data file had been confirmed, participants could delete the app and file from their phones.

4.3.2 Procedure

Data was collected from Monday-Friday from wake until sleep. Participants installed the mental workload ratings app on either the Sunday before data collection began or Monday morning, at their discretion. As mentioned, participants were notified for mental workload ratings every 30 minutes during specified working hours and every 1 hour outside of working hours. This was in order to aim to control for data entry intrusiveness and fatigue. Each evening, participants were asked to complete the set of questionnaires outlined above (PSS, revised POMS, VAS-F, and FCQ-S); participants were asked to complete the questionnaires around the same time each evening, which should be after their evening meal and close to their bedtime. The sleep diary was started on Tuesday and finished on

Saturday, in order to collect sleep data from Monday-Friday. All questionnaires were completed on Microsoft Word and instructions for how to input their answers was provided in the pre-study document. Participants could choose to send their questionnaire responses to the researcher after completion each day, or at once after they had all been completed. The post-study document was emailed to participants on the Friday.

4.3.3 Data Analysis

As mentioned, the analysis for the quantitative data was guided by the findings from the personal experiences section of the qualitative phase. Therefore, the methods for analysis are outlined after the qualitative personal experience results in Chapter 5.

4.4 The Qualitative Phase

4.4.1 Design and Procedure

The interview design was semi-structured and typically lasted between 1-2 hours. All interviews took place over Microsoft Teams or Zoom and were recorded. The semi-structured nature of the interviews meant that the researcher was guided by a set of pre-defined questions but participants were probed on individual topics that they mentioned and encouraged to talk at depth. Examples from participants' lives were encouraged whenever appropriate. When encouragement was needed, the interviewer used the participant's time in the quantitative phase as a prompt, by referring to a graph of their mental workload ratings throughout the week (Figure 4.1) and pointing out relevant sections that may help answer the questions; the diary data was also used as a prompt if needed, such as to ask how it felt when performing certain activities.

The pre-defined set of questions concerned 4 topics, and these were briefly outlined to participants.

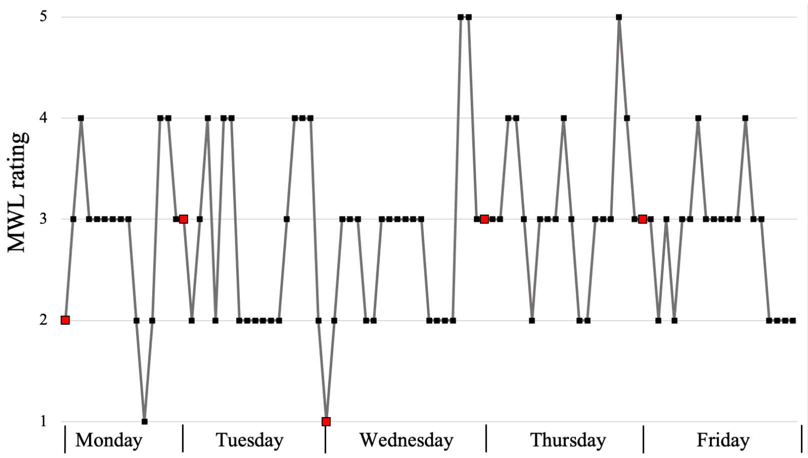


Figure 4.1: An example graph showing the subjective mental workload ratings given by P1 which was used as a prompt during the interview (red points indicate the first rating of each day).

Personal Experiences

The first section of the interviews regarded participants' personal experiences of mental workload (Appendix D). For each mental workload level (high, medium, and low), participants were asked to talk about their feelings, perceptions and attitudes about mental workload and the believed implications it has in their lives. For example, participants were asked questions such as how it felt to experience a high mental workload level, whether they are happy to maintain a high mental workload level, and whether they enjoyed periods of high mental workload.

Conceptualisation

The next section was more creative in nature (Appendix D). Participants were first asked for word associations for how it felt to experience the different mental workload levels (high, medium, and low) and for mental workload as a general concept. The researcher asked participants for 'a few' words that popped into their heads in association with the questions. Participants were then asked about any metaphors that they associated with mental workload and were asked (where appropriate) to make a sketch of these after the interview and send an image to the researcher. After outlining their own metaphor, the interviewer informed participants of other metaphors that had previously been manifested by prior research ([237]):

Others have used metaphors which describe mental workload as: on a spectrum that changes throughout the day, walking up a mountain where the steepness of the walk reflects the mental workload level, a thermometer, a brain filled with a set number of bubbles and the number of bubbles that pop depend on the mental workload level, and an input-output relationship where you put in a certain amount of mental workload and expect to get a return. Do any of these resonate with how you would think about mental workload?

Participants were then asked about the colours and shapes that they associated with their experience of being at the different mental workload levels.

Contextualisation

The third section was the shortest and simply related to *why* participants would like to track their mental workload data (Appendix D). The researcher provided participants with a hypothetical situation, stating:

Imagine that you owned an actual Fitbit for the brain – that is, a tracker that can objectively measure your mental workload levels in your life. You can reflect on that data by opening an app wherever and whenever you like.

Questions then regarded the reasons why participants would like to track this data (if they were in fact interested in tracking the data).

Ethical Considerations

The final section regarded participants' opinions about their ethical concerns and perceptions about the introduction and development of mental workload trackers (Appendix D). Participants were asked about data privacy, data sharing, and enforced mandatory tracking of mental workload data. For example, 'Would you be willing to share your data with people who might be interested in viewing it, such as your boss?'

4.4.2 Data Analysis

For the sections about personal experiences, contextualisation, and ethical considerations, an Interpretive Phenomenological Analysis (IPA) approach was used [207]. IPA is a qualitative approach that aims to understand how people make sense of their personal and social worlds in regard to their lived experiences and personal perceptions. [207]. Thus, the data from these three sections, strongly surround these aims.

IPA is modelled on people as self-reflective and self-interpretative beings who reflect on their experiences and try to interpret them [207]. In IPA, each participant's data is considered in depth to enable an idiographic approach before more general claims about the data are made [207]. Additionally, the use of IPA is especially suitable for topics that are contextual, subjective, and relatively under-studied [206]. IPA was favoured over thematic analysis for these sections because thematic analysis focuses on patterns across the data set; it does not provide a sense of contradictions within individual accounts, and the voices of individual participants can get lost [35]. Indeed, IPA does consider data patterns, but is also concerned with individual experiences [207].

A systematic approach for IPA has been outlined by Smith and Osborn [207], and has been adopted for this analysis. The interview data was transcribed verbatim. The lead researcher firstly familiarised themselves with the transcript. Comments were then noted in relation to first impressions and interpretations of the participant's account; different ink colours were used to indicate whether the comments were descriptive or interpretive. These notes were then translated into emergent codes. Once all emergent codes had been created, connections between them were identified and emergent themes were grouped together to materialise as initial subthemes umbrellaed under their superordinate themes. This was repeated for each participant, whilst using the themes from previous transcripts to orient the analysis. Respecting divergences as well as convergences in the data remained a priority throughout the analysis. After all transcripts had been analysed, a final set of superordinate themes and their subthemes were identified across the full set of data; the number of subthemes for each superordinate theme was reduced to only be representative of either rich or frequent data.

For the more creative conceptualisation section of the interview, IPA was not deemed as the most appropriate analysis method. This is because the data was less about understanding participants' lived experiences and deep personal perceptions, and more about identifying how mental workload data can be meaningfully communicated to a wide range of users. Thus, a thematic analysis approach [34] was used to analyse the interview data. This is a qualitative method which allows insight into collective or shared meanings or experiences by systematically identifying, organising, and offering insight into patterns of meaning across a dataset [34].

For the conceptualisation data, the data was firstly transcribed verbatim. The researcher familiarised themselves with each transcript, making notes on their initial interpretations or observations. For each transcript in turn, codes were generated which labelled portions of data in relation to our research questions. The transcripts and codes were handed to two other researchers (supervisor and colleague) for review to ensure credibility. All three researchers then worked collaboratively to identify themes from similar and overlapping codes. Themes were then reviewed to ensure they were representative of the codes and were supported by extracts from the transcripts. Any codes or themes not backed up by multiple data extracts were discarded.

Across the results section, each participant is referred to by a number, e.g. P15 refers to participant 15. The personal experience results are presented last as those findings guided the analysis for the quantitative phase, which are presented after the qualitative findings.

Quality Assurance and Positionality

To ensure good qualitative research practice, guidelines by Elliott et al [76] were followed. This involved verifying the credibility of the results by all researchers (supervisors) checking the data and collaboratively working with the data once transcribed. In particular, pairs of researchers discussed and challenged the emerging structure of themes and how subthemes related to each other, such that they went through several stages of refinement. Final themes were also subject to a team review, where the themes and their implications were presented and questioned. The final themes and data presented here are grounded in examples from participants to illustrate each theme and descriptive data about participants is also outlined. The perspective [76] and positionality [33] of the researchers are important to consider in qualitative research, as these are factors that can influence the research process [33]. This research falls within a WEIRD[2] context [149] represented by all researchers; 18 participants were UK based - 5 UK based participants were from South America and 1 participant was from and based in India. The researchers in this study all have a level of expertise pertaining to mental workload as a concept and it is reasonable to assume they have considered their own views on mental workload in daily life more than the usual office worker. Further, two of the research team self-described as hyper-organised, aiming to maximise their workload at work, and equally at home managing family life. The researchers recognise that their personal interests and assumptions about the topic may naturally play a role in their approach and understanding of the research outcomes [76].

[2]From their critique of HCI research: Western, Educated, Industrialized, Rich, and Democratic.

Chapter 5

Mental Workload for Personal Cognitive Informatics

This chapter aims to address research question (b):

Can a longitudinal and holistic approach to mental workload improve understanding of how mental workload could be valuable as a form of personal informatics, and mental workload as a concept itself?

5.1 Qualitative Contextualisation

To introduce the qualitative findings, this chapter begins by outlining the context for why participants were interested in objectively tracking their mental workload data in their daily lives as a form of personal cognitive informatics. Firstly, there was clear enthusiasm for tracking this data. The reason for this emerged to be for personal improvements, in terms of tracking for improved wellbeing and tracking for the optimisation of work habits.

5.1.1 Improved Wellbeing

Many participants reported their desire for the data to be used to improve their lives from a wellbeing perspective. Participant 13 provides an example of this:

"I guess I keep thinking about physical symptoms of mental workload – stress, pressure, feeling like less healthy and more stressed and stuff because you're under a higher mental workload. Like [the data could say] how much capacity you have to take on new stuff and deal with new stuff or, 'You need to just slow your mental workload down, lower your mental workload for a bit, recentre yourself' ... so you don't burn out so you can manage your work and personal life." (P13)

The passage above represents a common consensus amongst participants that the data could be useful for managing the affects that work can bring in terms of wellbeing. By adapting work based on the negative wellbeing effects that might result, such as burn out, the data was considered as having the potential to mitigate these factors.

5.1.2 Optimising Work Habits

As well as using the data to improve wellbeing, participants also frequently reported that they believed the data would be useful for improving their working habits, in terms of optimising productivity and results. Participant 11 described how they would use the data to improve their habits:

"I think I could use that information to change my habits of work, or like to be aware what I do, because it's something that I'm trying to change. I think we feel the more hours we are in front of our computer the better,

and I think that I've been realising lately, with the use of this app that I use on my phone, that sometimes I spend like twenty minutes on my computer being stressed and I don't do anything; I will just like waste my time on the internet and if I'm in those long periods with low levels of mental workload [and] I was not very productive, maybe I can think, 'Ok, what happened in that time?' And then I can think, 'Ok, what can I do to be more productive? Or to maybe work for less hours a day? Or work differently? Or try to experiment with other stuff.' So yeah, I think it will be awareness." (P11)

So Participant 11 reflects on how they would use awareness of their mental workload levels during particular times to improve their habits at work to increase their productivity and efficiency. Whilst we have outlined two examples of using the data for improving wellbeing and improving work, participants often reported that they believed the data could be used to improve both of these factors in their lives. Participant 2 provides an example of this:

"I think if you can look at an app or whatever and it lets you know that your brain says relaxed and it's most ready to do high levels of intense workload, for applying to things like work and stuff like that I think that would be interesting ... Is your brain relaxed? Is it feeling highly strained at the minute? And things like that, and then you use that read out to determine what you're gonna do next. So if it's at a low level of mental output then you might think, 'Ok, this is a perfect time to go and do that really intense rowing session,' or you know, 'Smash out a load of work,' or if you look at it and it's really high then you might wanna take some time to go for a walk or switch off or something like that. So I think having a current read out would be one thing and then using that to determine historic levels throughout the day ... and you'd then be able to tell, you know, 'We recommend that at, I dunno, half nine till half ten is the optimum time for

you to be doing your intense work, that's when your brain's most suited for it,' then it almost becomes a bit of a planner for you ... So knowing when you're going to be most productive but then also a wellbeing point of view. So it might tell you that you've been at a high level of mental activity for a long period of time, almost telling you that you need to stop basically, you need to take a break, so both productivity and also general wellbeing." (P2)

In Participant 2's rich account, they bring together both subthemes and describe the potential of the data collected to be useful for improving both wellbeing and working habits in terms of productivity and results.

5.2 Qualitative Personal Experiences

These findings were published in the Conference on Human Factors in Computing Systems (CHI):

Serena Midha, Max L Wilson, and Sarah Sharples. 2022. Lived Experiences of Mental Workload in Everyday Life. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22).

In terms of the interview section relating to personal experiences, four master themes were identified from the transcripts (Table 5.1): 1) general perceptions of mental workload, and 2) changing perceptions of mental workload, which together outline the fundamental perceptions of mental workload and the factors that can change these perceptions. Theme 3) the mental workload Cycle, is where we present a Cycle regarding the necessity to fluctuate between mental workload levels in certain patterns, and 4) the Cycle can't always be facilitated, outlines the factors that prevent these fluctuations.

Table 5.1: Table showing the final superordinate and subthemes from the Interpretive Phenomenological Analysis

Superordinate Theme	Subthemes
1) General Perceptions of Mental Workload *Describing the feelings associated with different levels of mental workload*	Positives high Negatives high Positives low Negatives low Positives medium
2) Changing Perceptions of Mental Workload *Describing factors that can affect perceptions of different mental workload levels*	Pressure Enjoyment Outcome Location
3) The Mental Workload Cycle *Describing how people use, combine, and manage the levels of mental workload.*	The cycle Sustainment is an issue Each level serves a purpose
4) The Cycle Can't Always be Facilitated *Describing factors that can disrupt access of different levels*	Life factors Internal factors External factors

5.2.1 General Perceptions of Mental Workload

Participants were probed about their general perceptions of low, medium, and high mental workload levels. For high and low mental workload, participants' perceptions were either positive or negative. Different perceptions at the same level (for high and low) were found among participants. For medium mental workload, participants' perceptions were rarely described negatively.

High Mental Workload

High mental workload was conceptualised by participants as either a state of deep concentration, or by a state of 'busyness' in terms of managing a large quantity of tasks. Indeed, some participants, such as P19, conceptualised high mental workload as having both of these dimensions: *"I would probably split it up into two where I gave high mental ratings I would say one where I was trying to do lots of different things at the same time ... but then there's the other side where there's a high mental workload where you are really focused on a particular activity and generally I perform better in that instance like if there's a high mental workload because I can focus on one thing and really dig deep and think about it."*

Alongside reflecting on what participants perceived high mental workload as 'being,' several positive and negative feelings were associated with operating at this level. Many participants described feeling fulfilled during and after periods of high mental workload: *"I think it's actually one of the things that in my opinion makes work a lot of times rewarding you know being able to think hard about stuff"* (P5). Similarly, P9 said *"I think when you solve things that are challenging it's like you feel comfort at the end and you feel like you did something like well and I feel relaxed after the day if I say that I had really high mental workload but I was able to overcome it."* Participants also felt enjoyment, positively stimulated, and less distracted (which were each recorded as codes in the analysis) and thus it is clear that positive associations are often made with being at a high mental workload level.

Significant negative feelings were also associated with being at a high mental workload level. Participants often described high mental workload as feeling like pressure, and could be experienced as stressful and sometimes overwhelming: *"I just feel like stressed and I know that I need to prioritise and maybe I'm struggling to prioritise at that time cause I feel like there are too many tasks that I need to look at the same time"* (P6). *"I never really long to be at a five [the highest mental workload rating]. I think it feels quite out of control being at a five for any long period of time"* (P16). As a result of these negative feelings, participants may avoid operating at that level. Taken together, high mental workload can be perceived at opposite ends of the spectrum in terms of positive and negative feeling associations, and this is individual to each person. As maybe to be expected, this was also true for feelings associated with being at a low mental workload level.

Low Mental Workload

Low mental workload was typically described as the feeling where one can operate on autopilot, or as a state where one feels like there is a lack of activity (whether good or bad). *"I wanna say it's just sort of like the default feeling I guess nothing interesting's really happening it's just sort of I wanna say sort of tedious"* (P15). P14 described it as when everything has been achieved for the day: *"See I think that [low mental workload] counts as when you've not got anything to do so like at 10 o clock when the kids are like in their bed that's when I would say I have a low mental workload because I've not got anything else to do."*

These two conceptualisations of low mental workload appear related as they are both associated with low levels of demand, and like the associations with high mental workload, operating at a low mental workload level was generally associated with both positive and negative feelings. Participants positively described low mental workload as relaxing, enjoyable, and indeed more manageable: *"[A low mental workload day] it'd be an enjoyable chilled day just recharging chilling enjoying yourself"* (P13). Perhaps more surprisingly, some participants reported low mental workload had an impact on how they view the world around them in terms of manageability: *"If I'm just bumbling along I feel like my whole life feels a bit more in order like personally and at work like it feels a bit more manageable."* (P13)

Participants also reported a considerable amount of negative feelings generally associated with being at a low mental workload level, being boring at the least. *"When you're operating at a low level it becomes quite mundane. Everyday just feels a little bit the same, I suppose that's how it is"* (P16). Many participants felt unsatisfied and unproductive at a low mental workload level: *"Oh my gosh I would try to fill it in with anything ... just to feel like 'Ok I did something a little bit productive today'."* (P7)

Participants often reported feeling more prone to distraction when at a low mental workload level: *"I have a lot of distractions and I look at my phone like too many times on the social media like Instagram, Snapchat or Facebook. Even though I know that I have no messages I'm just opening them, I see I have nothing and then I just close the apps"* (P8). This supports related research that showed people self reporting as more easily distracted were more likely to be so at low mental workload [92]. As specific codes, the feelings of boredom, unproductivity, dissatisfaction, a lack of enjoyment, and distractable were negative feelings generally associated with experiencing a low mental workload level.

It is worth noting that there was no consistent form of polar trend between perceptions of high mental workload and perceptions of low mental workload; for example, participants that generally had positive feelings associated with high mental workload did not necessarily have negative feelings associated with low mental workload.

Medium Mental Workload

Whilst high and low mental workload levels were associated with both positive and negative feelings, medium mental workload was overwhelmingly associated with positive feelings. Medium mental workload seemed to be perceived as ‘a happy medium’ which counteracted the negative associations of the high and low levels, whilst retaining in some form their positive associations: *“You’re in danger in the low to procrastinate on stuff to other things because you don’t feel mentally challenged and then conversely on the high you might wanna try and avoid it because of the taxation of it. I think medium’s a sweet spot if that makes sense where you know that you’re using a bit of mental load but you’re not overdoing it”* (P19). Here Participant 19 described medium mental workload as having the right balance of intensity and activity; given their earlier account of high mental workload, it can be assumed that this passage applied to activities related to either quantity or mental absorption. The association of medium mental workload as being the ‘sweet spot’ was reflected by other participants who further disclosed that the balance between low and high levels generated feelings of comfort, enjoyment, and control. E.g: *“It’s comfortable it’s like you are not at your high level of stress or things to do but you are not without doing anything I think it’s cool to be there.”* (P9)

5.2.2 Changing Perceptions of Mental Workload

Where the previous theme outlines underlying perceptions of mental workload at each level, this theme outlines certain caveats (pressure, enjoyment, outcome, and location) that can change these perceptions.

Pressure

Participants reported that high mental workload tasks that were associated with pressure, in the form of external pressure or time pressure, resulted in negative feelings towards the task: *“If I’ve got a busy workload and it’s not* *like super intense deadlines I don't mind that it kind of makes time pass quicker, [and you've] got things to focus on but yeah if it's like I have to do everything now I don't enjoy that at all. I don't think I deal very well with stress and intense pressure so I don't think I enjoy that"* (P13). In this case, Participant 13's experience of a high mental workload level is turned more stressful by the time pressure associated with the task, to the extent where a fairly positive experience of high mental workload is turned into a negative one that is associated with stress and pressure. Indeed, when stress was mentioned in interviews in relation to mental workload, it was typically in relation to immediate or consequential pressure.

Enjoyment

Whilst pressure is a reported factor that seemingly affects perceptions of high mental workload, task enjoyment was described as a factor that affected the perceived feelings of all mental workload levels: *"If it's something like work or whatever home task that requires a medium mental workload or if it's something like sports or playing a competitive game then it definitely feels different, probably to do with the enjoyment that's associated with it"* (P15). So enjoyment of the task or activity affected the experience of being at a medium mental workload level for Participant 15. This caveat did not necessarily mean that Participant 15's medium mental workload experience went from negative to positive when participating in an activity that they associated with more enjoyment, as they reported simply the change in experience between activities.

Outcome

Perceived experiences of different mental workload levels were often found to be affected by the outcome of the task or activity, particularly as a factor associated with the negative low mental workload experiences reported above. Participant 8, however, reflects on this in terms of amount of work produced from high mental workload periods: *"If I'm being productive I do [enjoy high mental workload] ... but if I feel like I haven't advanced or progressed that much I feel stressed cause I feel like I'm wasting my time, I have no good results"* (P8). So Participant 8's whole experience of being at a high mental workload is dependent on whether they are progressing through the task(s) at a satisfying rate, and thus the outcome, or ongoing outcome, of the task is a key factor in the perception of mental workload. This finding was also identified for the low and medium mental workload levels.

However, it is not only personal assessment of the task output that might affect how the mental workload level is perceived. Participant 1 described how the external response to the output can influence how the level is perceived: *"So I feel better in meetings when there's some kind of positive feedback of some sort ... When I talk my eyes wander a lot or I look around my room cause there's not a person to look at, and when I look back, some calls I just see a bunch of blank faces staring at screens or like at their own thing, and sometimes I look back and they're smiling and nodding and I'm like, 'I'm doing alright, the point I said's valid' and those ones feel better at the same workload. So I'm trying just as hard to make a point and there's like an extra good feeling that comes from looking back and someone's smiling or nodding and like 'ok that was a good point'"* (P1). This passage indicates that not only does the internal assessment of the output affect the perception of the mental workload level, but the internal assessment of the external response affects the perception of the mental workload as well.

Location

The environment in which a low mental workload level is experienced was a recurring code that affected how participants perceived being at a low mental workload level: *"It feels less guilty when I'm outside work because when I'm doing work and I'm rating myself low it almost feels like it means I'm not doing much or I'm not doing enough and I think it brings about some sort of guilt which is weird because it shouldn't but it still does"* (P5). As an interesting example of reflection on this kind of data, Participant 5 describes how being at a low mental workload level at work (as a Research Fellow) generates feelings of guilt which is not present when outside of work hours. They acknowledge that there is no basis for that association, but reflect on how it almost cannot be prevented in that environment.

Location was also often reported to affect the enjoyment factor, above, of being at a low mental workload level: *"If it's [low mental workload] at work I become apprehensive, maybe a bit irritated and I am anticipating the boredom. But if it's personal life then yeah probably quite happy [to be at a low mental workload], so I'll spend the day doing the cleaning washing the pots doing some laundry watching the telly going for a walk all very low mental workload stuff but I'm quite happy to do that"* (P10). As reported in several places so far, it is evident that some people alter their priorities for low and high mental workload for different parts of their life, as well as within parts of their work; Participant 10 has negative associations of low mental workload when working (as a Chartered Accountant) but reported more enjoyment of being at a low mental workload level in their personal life.

Summary

Taking the two themes presented above together, we can reflect on the different perceptions at the same level for high and low mental workload. The positive feelings high mental workload was associated with included feeling: fulfilled, enjoyment, stimulated, and less distracted, whereas the negative feelings included feeling: pressured, stressed, and overwhelmed. The positive feelings associated with low mental workload included feeling: relaxed, enjoyment, and manageable, and the negative feelings included feeling: bored, distracted, unsatisfied, and unproductive. Medium mental workload was only perceived positively, as a "happy medium," in terms of feeling comfort, enjoyment, and in control. These initial perceptions are subject to change, however, with the presence of caveats (pressure, enjoyment, outcome, and location) which can completely change the mental workload perceived experiences.

5.2.3 The Mental Workload Cycle

Theme 3 presents an apparent Cycle in which fluctuations between mental workload levels are important for increased wellbeing, optimal performance, and positive mental workload perceptions. This is because each mental workload level serves a different and important purpose to the individual, and negative consequences are likely to happen if any level is sustained for too long.

The Mental Workload Cycle

The Cycle describes how participants fluctuated between the different mental workload levels in specific *patterns*. *"Whether it's low, medium, or high, you have to have variety to be the most efficient person. If I was to define a perfect day it would be a mix, so some low mental tasks, some medium mental tasks, some high mental tasks. That's kind of the days I'd probably define myself most efficient"* (P19). P19, a senior health economics manager, described fluctuating between mental workload levels in terms of efficiency; they feel like it's the balance between the levels that enables an efficient self. What the passage echoes from many participants, is that they deliberately seek out fluctuations in mental workload levels, often in particular patterns:

Firstly, after experiencing a high mental workload level, participants would typically transition directly to a low mental workload level: *"I do seek out low mental workload breaks so if I'm doing something that's going to take me a long time but is at a sustained high level, probably every couple of hours I will go and look at my emails and just reply to a few things and then come back to it"* (P10). While P10 described the fluctuations within tasks as a break, some described transitioning to longer periods of low mental workload after longer periods of high mental workload had been completed: *"I have a band and we sometimes record in a studio. We have to do it for the whole day because I mean it's hard to get people into the same room on the same day so we have to go there from like I dunno, 9am in the morning to 9pm or 7pm in the evening. So it's constantly high mental workload, listening and getting comments and feedback and everything ... After I'm done I'm just gonna chill, just find something that really disconnects me, like reality TV or something like that."* (P7)

Next, participants actively sought out higher mental workload levels when experiencing a low mental workload level: *"I would seek high mental workload [when at a low mental workload level] . . . I often have a lot to things on my to do list so I can create high mental workload by not doing things in my own mind cause they're there playing on my mind, so by doing them, A, I'll get some reinforcement out of whatever it is, but it can dial down that anxiety about not getting things done"* (P18). Similarly, Participant 1 described that they seek to raise their mental workload level to seemingly anything above a low level: *"I could sit in front of the TV all day not really doing much, flicking through my phone or watching YouTube or whatever . . . but it doesn't feel very good... If it's in my power I'll put tasks in there to make it higher. It's the reason that I book meetings in or find new opportunities I guess... I'll put something in on purpose to stimulate myself"* (P1). Whilst Participant 1, like many of the participants, did not specify which mental workload level they would transition to from a low mental workload level, it is clear that they aim to transition to a 'higher' level. Further passages described activities that are undertaken after experiencing a low mental workload level, we interpret that participants did seek out medium mental workload levels as well as high mental workload levels after a period of low mental workload: *"I do yeah most definitely [seek higher mental workload levels when experiencing low mental workload]. I've always been a big reader, always read a lot of books and they're not always, you know, highly cerebral or anything, they are trashy novels quite often, but just to keep the brain working I got through about 40 different books last year so I will always seek out something."* (P10)

Whilst the general consensus in theme one was that participants were more happy sustaining a medium mental workload level compared to a high or low level, participants did eventually seek out either a high or low level.

Interestingly, each participant had a clear preference about whether they would seek up or down from a medium mental workload level. Participant 14, for example, tended to seek out a high mental workload from a medium mental workload level: *"I'd probably be quite happy there [at a medium mental workload level] but I would probably always tend to seek for the higher workload rather than the lower one"* (P14). On the other hand, Participant 13 preferred not to transition to a high mental workload level from a medium level; instead they would seek for a low level of mental workload. *"I think I'm happy to stay at a medium mental workload for a relatively long period of time. I think eventually you probably would seek out low levels, I don't think I'd ever feel like I need to seek out high levels."* (P13)

From the findings in this subtheme, participants aim to fluctuate between mental workload levels and this occurs in specific patterns. Specifically, from a high mental workload level, participants seek out a transition to a low mental workload level. From a low mental workload level, participants seek out a higher mental workload level. And from a medium mental workload level, preference for transitioning to a high or low mental workload level varied by participant, but each participant had their own preference.

Sustainment is an Issue

It seemed as though a reason that participants sought fluctuations between mental workload levels was driven by experiences of sustaining any mental workload level for 'too long,' which resulted in negative consequences. These related to wellbeing, work output, mood, and perceptions. Burnout, for example, was a commonly reported consequence: *"Up until very recently I would literally say ... 'this is what I want to achieve at the end, this is* *what I need to do to get there' and then I'll just do it regardless of whether I've worked a long day ... But then I think that's ended up leading me to get burnt out before so actually since I've be furloughed, I'm trying to aim for more of a balance. So for instance on that Thursday [during the previous week] when I had a really hectic morning, or a really intense morning, I just decided to put everything away and just went out for a really casual jog ... It was quite a good way to switch off"* (P2). The passage above from Participant 2 is an interesting reflection on the consequence of prolonged mental workload and other experiences in their life, and the kind of goal forming that we think may develop in more detail with future wearable technology. Now, P2 purposefully inputs periods of low mental workload as breaks in order to counteract the negative effects they have recognised.

Sustaining a high mental workload level was also commonly associated with feelings of fatigue: *"I remember a couple of weeks ago there's one day where I was really focused on something, and I think it was a four hour meeting that I was in, and I had to be on the ball all the way through that four hour meeting, and I remember at eight o clock that night sitting there and going 'I need to go to bed,' because I'd just kind of completely gone"* (P19). After sustained periods of high mental workload, many participants reported feeling more likely to put off non-essential life tasks, such as the washing up: *"Things like cooking dinner I suppose you have to do it, you have to just get on with it but I will, if I've had like a really long day and then it's been like a long evening, I'd probably just be like 'No I'll leave the dishes and I'll leave the washing' things like that I do have less motivation to do [after sustained periods of high mental workload]."* (P14)

Participants reported experiencing negative moods as the product of sustained high mental workload. Participant 17, for example, said: *"I might not be at my best behaviour with others. I don't vent out but still I don't* *reply quite politely or if anyone asks multiple things I get irritated, so that is a downside of high mental workload."*

Participants also reported resentment towards their work, or a loss in the quality of work output as a consequence of sustaining a high mental workload level; this consequence is different to the other consequences outlined so far in the sense that its affects take place during the high mental workload level itself rather than as a wider implication. Participant 12 described both of these consequences: *"With copywriting I love it but I feel if I just spend ridiculous amounts of time doing it without a break I'd become almost, not detested, but I wouldn't feel as passionate or as loving towards it, for want of a better word. And after a certain amount of time, I think my output and quality of work would definitely decrease as well due to not having that rest or time away from the screen to focus."* (P12)

Some participants even reported physical health consequences from sustaining a high mental workload level: *"Last week I had a really urgent deadline and because I knew it had to be with a client by the end of the day, I was working super efficiently and the director was doing it alongside me, but at the end of the day I was just dead, like exhausted, had a headache ... after feeling the pressure all day by the end of it yeah I had a headache and I kinda felt a bit spaced cause I was just focused on one thing all day and it was very intense and then yeah, I really think the headache was really related to concentrating on one thing so solidly all day"* (P13). So by the end of Participant 13's period of sustained high mental workload, they were not only feeling fatigued, but also suffering from a headache and not functioning to their perceived normal level. This really captures the type of physical health consequences that sustaining a high mental workload can have. Not only were physical health consequences reported, but Participant 16 related sustained high mental workload with mental health consequences: *"I don't think it's something [being at a high mental workload level] that's sustainable for huge periods of time ... I don't think it's something that's particularly healthy for long periods of time . . . I think people can really suffer in terms of their mental health when you're operating at such a high level for a long time ... It's such a high pressure because you're just operating at a level where you're just waiting for something to go wrong."* It appears P16 associates high mental workload with pressure (see theme two), and sustaining the pressure is the factor which can have negative mental health consequences for them.

Sustaining *low* mental workload levels also had negative consequences for our participants in terms of how they perceived their experiences of the level. Participant 2 reflected on their recent experiences of how their perceptions of mental workload were shaped by sustaining a low mental workload level for too long: *"Whilst I was furloughed basically everything was just a low mental workload, I didn't really have much to do, and I don't find that enjoyable because I feel like you're not achieving something or like there's not really not much purpose to it. But then since I've started working again and since I've started training more again, I think when you have lower periods balanced with higher periods it makes the lower periods more enjoyable, more relaxing, cause you've actually got something to relax from and almost they feel like earnt or deserved . . . I think you need to have the highs and the lows to enjoy both and I don't think life would be rewarding or enjoyable if you are constantly sat at either end of the spectrum"* (P2). Participant 2, while furloughed, found themselves operating at a low mental workload level constantly, and their experience of that level was perceived negatively in terms of unproductivity and dissatisfaction. When their normal activities resumed again, and their daily mental workload levels were varied, the low mental workload experience was perceived in a much more positive way. Participant 16 reported a similar experience: *"I think I can enjoy it [low mental workload] when I can put it in context to a high mental workload. I think for me it's a bit like if every day was Christmas I wouldn't enjoy Christmas"* (P16)[1]. Participant 16 was going through a quiet period at work (as a Programme Support Officer) which predominantly consisted of low mental workload levels. They describe how without the balance of high mental workload, low mental workload loses its enjoyability (see theme two). Thus, both of these passages highlight that the low mental workload feeling can be influenced by how long the level is sustained for.

In terms of medium mental workload, while it was considered as the most sustainable level (see theme one), sustaining it for 'too' long still left some participants missing the full level of excitement which was associated with operating at a high mental workload level: *"Being at that medium is good, but it's even better when you've got the context of the thrill of sometimes having to be at that greater capacity"* (P16). Participants who sought out low mental workload from medium mental workload (as described in The mental workload Cycle) seemed to require low mental workload levels as a break, suggesting that sustaining a medium mental workload level is still fatiguing: *"I feel like medium mental workload you're kind of balanced but eventually I'd be like, 'yeah I just need a little break'."* (P13) However, the reasons for seeking out a low mental workload level from a medium mental workload level were not fully established, and thus issues with sustaining a medium mental workload level were not revealed in detail within our data.

To summarise this subtheme, sustaining any mental workload level for too long resulted in negative consequences. These related to wellbeing, work

[1]For clarity, in this case Christmas was being referred to as a special occasion, which wouldnt be special if it was regular, as opposed to referring to Christmas as being specifically high workload or low workload

output, mood, and perceptions. Specifically, sustaining high mental workload for too long was associated with: burnout, fatigue, negative mood, increased resentment, reduced work quality, and decreased physical and mental health. Sustaining low mental workload was associated with: decreased enjoyment, decreased productivity, and decreased satisfaction. Sustaining medium mental workload was associated with a lack of excitement and potential fatigue.

Each Level Serves a Purpose

The final subtheme outlines the reasons why each mental workload level is important to include in the mental workload Cycle. We found that each level of mental workload serves a different purpose.

High mental workload was related to positive implications for work and internal perceptions. Many participants associated high mental workload with increased work output: *"I feel the most productive, I get more things done. So like times when we used to go into the office, it sounds like a long time ago, I could go in and if I had like three hours of really high mental workload, I could be really productive [and] I could come back home by lunch because I finished what I wanted for that day"* (P7). As well as speed of output, quality of output was also associated with high mental workload in our participants: *"Often by operating at a high level of demand on yourself I found the pace of it brings a greater quality in your work that isn't there when you're operating at a two or three as well. I find that demand often spurs me into doing some really great pieces of work."* (P16)

Perhaps more important than 'better' output, high mental workload was also associated with harder tasks that cannot be completed at a lower level of mental workload: *"I think when I'm at a medium mental workload it* *refreshes your brain enough to feel up for taking the plunge with some higher mental workload tasks. I think sometimes you go, 'Alright, I am going to find some time to do this other thing,' because you're not being overloaded by loads of stuff, you can start to think again a bit more creatively which, in some instances you can do in low mental workload as well, but I think it's quite a nice feeling to feel like you're enthused about doing some harder things"* (P18). From the passage above, Participant 18, along with several participants, associated high mental workload with challenging tasks, and indeed sense of achievement: *"I think if you've done something in the high space generally if you were to evaluate at the end of the day, if you've managed to achieve something when you've had a high mental workload, generally you feel more exhausted, but you almost feel more happy that you've managed to achieve something which is generally quite taxing. So, I'd probably say there's more of a degree of self worth at that high mental workload element"* (P19). Feeling that sense of achievement was almost like an indirect effect of high mental workload for many, as the high level is associated with taxing output which is then associated with a sense of achievement if the task is completed satisfactorily.

We also saw that low mental workload is an important level to obtain for a number of reasons. Firstly, for our participants, low mental workload was important for a mental rest and recovery: *"Now that I'm working at home I just make sure that I build in tea breaks and loo breaks and things like that so I can refresh"* (P10). Participant 10 describes how they actively ensure there are periods of low mental workload in order for them to refresh and recharge before entering higher mental workload levels again. Equally, participants often reported that low mental workload could be used in preparation for a high mental workload level: *"There are days like last Thursday where I need to go home without anything to do so I can* *use that day to be more relaxed and get energy for the next one"* (P9). Participant 12 described these breaks, in particular, as 'clearing the mind': *"I do definitely try and seek out tasks or moments where I can just lower my mental workload level . . . If I do have a break the first key thing is to just get away from my screen, that mental disconnection from work, from being in front of the computer, from being sat down in a set position looking at the same windows and walls. Yeah, just even just going outside to grab some fresh air, it gives me time to just come back with a different perspective or if I'm struggling to get motivated or get my workload at a decent output, just coming back with a different perspective, different mindset ... Recharge the batteries."*

In many cases, low mental workload was used as a reward: *"Something I've noticed [during] lockdown, I kind of set myself one or two large tasks each day, or a large task, and if that's done, I'll just chill out rather than trying to get loads more"* (P1). This is important given that reward-based behaviour is often intrinsically motivating for some people.

For medium mental workload, the level served to balance the characteristics of the low and high levels. It was regarded as the most sustainable level and had positive implications in terms of productivity and personal perceptions; Participant 6 sums up the implications of medium mental workload: *"I think it's the best place to be from a personal and a productivity point of view I guess."* (P6)

In summary of this subtheme, it appears that different mental workload levels do different and important jobs. High mental workload was associated with increased work output quantity and quality, harder tasks, and a sense of achievement. Low mental workload was associated with rest and recovery, preparation, and reward. Medium mental workload was associated with productivity and positive perceptions.

Summary

In the theme presented, our findings show that there is a Cycle for mental workload (Figure 5.1), where individuals require mental workload fluctuations in particular patterns. There are negative consequences of sustaining levels for 'too' long (particularly low and high levels), and there are reasons why each level is important to incorporate into the cycle. This theme furthers our understanding of what type of goals we should aim for in terms of our mental workload lives, as identifying the cycle right for each individual could result in a more sustainable, efficient, and satisfying way of living and working.

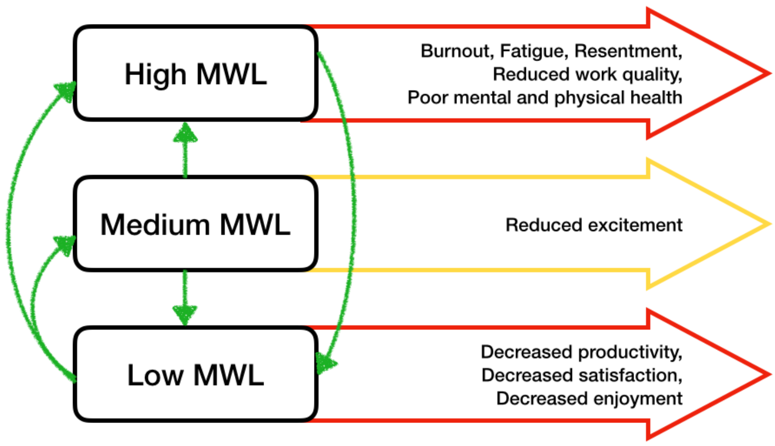


Figure 5.1: Diagram showing the apparent Mental Workload Cycle experienced by participants, and the consequences of sustaining each level for too long.

5.2.4 The Cycle Can't Always be Facilitated

We have outlined that fluctuations between mental workload levels are important, but it appears that they are not always possible. Life, internal, and external factors were found to interfere with achieving the required, or desired, mental workload levels. This was typically by either causing participants to remain operating at a certain level, or preventing them from achieving desired changes.

Life Factors

Two participants provided particularly rich accounts of how their perceived ability of mental workload was affected by medical factors. Participant 18, a Voluntary Services Deputy Manager, reflected on how they felt that their capacity for mental workload was decreased: *"I started IVF treatment last week so it was really interesting for me that my capacity for mental workload went through the floor. So you'll see I had a few really bad days [in the subjective ratings], but there was nothing bad that happened on those days that's out of the ordinary for my work. There were peaks and troughs in the mental workload coming in, but when there was a small peak, to me it was a massive peak and my brain just went, 'Ahhhh.' Last week, might, in hindsight, not have been the best week for me to study for you because it was very visceral, it was very physical, the feeling of complete overwhelm that came with like three people asking me things at once, which I would normally need a lot more things for that to happen to me I think. Yeah, pile hormones into your body, who knew it would change your personality a bit ((laughs))"* (P18). From Participant 18's passage, we see that undergoing hormonal medical treatment seemed to implicate their experience of operating at 'normal' mental workload levels. They found that tasks which would normally require lower levels of mental workload instead were experienced as high mental workload, which felt overwhelming. On the other hand, Participant 3 did seem to have perceived control over their mental workload levels, but purposely avoided operating at a high mental workload level due to exacerbated negative consequences: *"I'm on long term medication and the main side effect is somnolence, which is tiredness, but not a feeling that you're about to fall asleep. I've lived with it long enough now that I just take it in my stride, I don't feel like I change anything day to day, but I have had to adapt and I think that might explain one of the reasons why I avoid fives [the highest mental workload level] because I'll just get exhausted. I aim for three [medium mental workload] and I'm happy with a three, because I know I can sustain it. In my previous job, when I was probably under a lot more periods of four [high mental workload], I had to come home midweek and sleep for the entire Wednesday afternoon ... It's a permanent state of being slightly subdued and I think it has been really interesting to see for me this week how much I seem to, without having realised it, maybe I adapt my workload levels to what I feel I can actually achieve and sustain"* (P3). So Participant 3 described their perceived inability to operate at high mental workload levels due to the side effects of prescribed medication; if they did reach a high mental workload level, they feared the consequences that would have on their life.

Some participants reported that exercise affects their perceived ability to operate at a high mental workload level: *"I do tend to find in the mornings if I do go for a run prior to work my overall output and mental capacity for workload is a lot higher. I think that I've started the day off well and I've set out to achieve something and I've done it, so I seem to almost take that into the working environment"* (P12). While exercise is kind of a pre-workload activity, it also appears that e.g. music, as a concurrent context, might have some effect on some participants' perceived ability of reaching certain mental workload levels: *"I play music to help me to concentrate"* (P8). With a small number of comments about the use of music to facilitate achieving mental workload levels, this might be a factor that could be explored further.

Internal Factors

As well as life factors affecting the obtainment of certain mental workload levels, internal factors were found to have the same effect. Participants often reported that, while aiming for low mental workload, their level was elevated by internal thoughts which were unrelated to the activity they were doing: *"I'll go to sleep, well not go to sleep, I'll think about work in bed which is always annoying, and particularly when you have very transactional work cause it's not always one story line that you're thinking about. You're not thinking about a project, you're thinking about, 'Do this, do that, do this, do that,' and that's really annoying cause it's like a cacophony of thoughts all at once, it's just not conducive to sleep ... Literally just in my head whilst I'm in bed I have a mental workload which is ridiculous."* (P18) An increase in mental workload level due to internal thoughts was reported by many participants, and was shown affect obtaining certain mental workload levels.

Another factor that participants often reported as being a barrier to achieving a high mental workload level was that reaching a high level requires effort: *"When I have high mental workload I have to be fully concentrated, so it's something that I have to plan and something more about will power ... Sometimes it's difficult to get to that and I find it difficult to concentrate, I get distracted"* (P11). So Participant 11 reflected on how achieving a high mental workload level takes effort in terms of planning and internal will power. Sometimes, they struggle to reach the high mental workload level, even though they try. Enjoyment of the task was often the factor that affected whether this effort barrier was easily overcome: *" If I'm en-*

joying it [a high mental workload task] it's like intrinsic motivation ... I do enjoy some PhD work and that's the stuff I'll keep doing, the stuff's that's sustained longer is because I've enjoyed it" (P1). For future work, we could speculate that enjoyment could result in either not maintaining mental workload levels for long enough if it is a task that they do not enjoy, or possibly maintaining certain mental workload levels for too long if it a task they do enjoy, both affecting the balance of the cycle.

External Factors

As well as life and internal factors potentially affecting the mental workload Cycle, external factors also appear to interfere with the balance of levels. External demands were reported by participants which required sustaining mental workload levels: *"I had to do it [maintain a high mental workload level] in the lab sometimes, it's like you go there [at] nine o clock in the morning and you can't go home till five in the afternoon without lunch because you cannot turn off the reactor."* (P9) From Participant 9's passage, it appears that they are sometimes placed in a situation where because of the demands of the task, a high mental workload level must be maintained, which does not facilitate the balance of mental workload levels.

As initially reported in theme two, pressures were reported to result in sustaining mental workload levels rather than fluctuating. Participant 19 noted that a period of leave from work left them feeling unable to avoid sustaining a high mental workload level when they returned. *"Sometimes you can't avoid it [sustaining a high mental workload level] like, perfect example, so I'm on holiday from Sunday to Wednesday this week and I'm back Wednesday night. I've got really important meetings Thursday and Friday, external meetings and internal meetings, my diary is full. I couldn't* *really get away from that because I'm away Monday to Wednesday."* (P19)

Similarly, many participants reporting having to sustain high mental workload levels simply until their task was completed (regardless of cause): *"I'm happy to maintain those [high mental workload levels] until it's done. Usually I do allow myself buffer time, for example if I think that this specific section is going to take me two days, in my mind if I enter in a high mental workload area then I finish in one day or half a day, I'm like, 'Ok this is good, I'm fine with it,' so I believe it's more goal orientated than time orientated in my case"* (P7). Thus it seems that Participant 7 was willing to maintain a high mental workload level for as long as it took to produce what they perceived as a satisfactory amount of output (as discussed in theme two). We could speculate again that this could result in sustaining a certain mental workload level for too short or too long for the individual. It is worth noting that sustaining mental workload levels instead of fluctuating between mental workload levels does not only apply to high mental workload with our participants; for example, as previously mentioned, Participant 2 had a period of sustaining low mental workload levels because of external factors as they had be furloughed from work and their sporting activities had been paused.

Summary

Whilst participants benefited from fluctuating between mental workload levels, this was not always possible. Life factors, including medical reasons, exercise, and potentially music, internal factors, including thoughts and effort, and external factors, including circumstances and task completion, often interfered with fluctuations. This was likely to result in the negative consequences described in theme 3 as levels were either sustained or not achieved.

5.3 Quantitative Data Analysis

This section will outline the decisions made about the data collected and the methods for analysis, in terms of how the data was handled and the statistical tests used. The results section will then adopt more of a commentary style; it will briefly provide each part of the analysis with context in terms of how it relates to aspects of the qualitative findings, and then outline the specific analysis that was done in relation to this. This was intended to benefit the flow and clarity of the results by limiting the need to refer back to the previous section in order to understand the results.

The vast amount of data collected was firstly reviewed to determine what data could be useful to further develop the qualitative findings. The variety and quantity of data could be used and analysed in countless directions, so following an analysis plan that was within scope of the qualitative findings was a priority.

In terms of the ratings app, which collected responses to 4 questions, it was decided to not go forward with questions 1 ('What is your current level of mental workload?') and 3 ('What would you rate your overall performance since the last rating?'), and only consider data from questions 2 ('What has you overall mental workload level been since the last rating?') and 4 ('Please use the space below to report a summary of the tasks performed since the last rating.'). This was because considering 'current' mental workload levels could have captured data of largely varying lengths, such as a few seconds to several minutes (up to an hour), and the analysis would not have been able to differentiate between this potentially critical difference.

Thus, that data did not seem useful for understanding the fluctuations that people have in their days or the impact that mental workload has on their lives. Additionally, evaluating subjective performance was considered out of scope for this analysis as it was not closely related to the qualitative findings or thesis research questions. Thus, reviewing subjective data in isolation from questions 2 and 4 was the initial point of data exploration.

Scope of Analysis

Based on the qualitative findings from the personal experiences section, there were two points of investigation that our quantitative data could build upon. The first one involved the patterns of fluctuations that participants had between different mental workload levels, which could be explored by the data collected from the ratings app. The second involved investigating the impact that mental workload had on other factors, including evening questionnaire data (e.g. fatigue levels) and phone usage data.

A challenge for this second area of investigation that explored the relationships between variables, was the different time intervals between the measures. The mental workload ratings data was sampled multiple times each day (approximately every 30 minutes or 1 hour), data was collected once a day for the questionnaires, and every hour for phone usage data. We therefore needed a way to make connections between the different types of data. To do this, we aimed to quantify the mental workload data into values that would represent a certain amount of time. For example, a 'mental workload day' would consist of one value that represented the type of mental workload day a participant had experienced, and would enable the data to be explored in relation to the evening questionnaire data.

A time-weighted average was selected for this purpose of quantifying multiple mental workload ratings into one value. Time-weighted averages are a way of getting an unbiased average when there is irregularly sampled data. Our data is irregularly sampled due to the 30 min rating requirement during working hours compared to the 1 hour requirement outside of working hours. Additionally, participants did not always rate as soon as they were prompted; there were large variations in times between responses for all participants. A typical average does not account for time, and the data is therefore distorted as more frequent ratings would carry more weight.

A weighted average was therefore used to account for time. It was calculated by taking the area under the curve and dividing it by total time. This way, each value is given a weight depending on how much time it represents. The longer the time interval, the bigger the weight. Mixed models were then used to investigate the relationship between mental workload levels and the other factors, outlined below.

Mixed Model Analysis:

The purpose of linear mixed models is the same as typical linear models such that the factors that are having an impact on an outcome variable can be investigated. Typical linear models have an assumption of independence of data points, such that they are not suitable for repeated measures designs. Linear mixed models are an extension of typical linear models, such as linear regression, and are often used with more complicated sampling designs. As well as fixed effects, which are the variables of interest, mixed models include random effects. Random effects are effects that can create random variance in the sample. For example, in a repeated measures design there are multiple data points from the same participant. This means that observations are nested in subgroups that could vary by sample. Random effects therefore account for certain factors that could explain some of the random variance in the sample.

To estimate the value of an outcome variable from the predictor variable, a standard regression line could be plotted, which places the intercept at the population mean. However, the data points in the sample may be far away from the plotted line because it does not take into account the differences that random effects might have on the data. Mixed models, however, take into account these random effects by fitting a regression line for each random effect, e.g. each participant. These lines likely have different intercepts as each participant will have a different value of the outcome variable. The slope of the line is often fixed, with the researcher making the assumption that the population has the same slope trend, and the slope is therefore estimated from the slopes of the sample data points. The different intercept values detail the distance each intercept is from the mean intercept. Because each participant has their own regression line, the error of the model fit will be low, and the error residuals should be uncorrelated to each other.

A mixed model can therefore be performed, which includes the random variance in the data caused by the random effects, as well as the random variance from the error term that is already accounted for in typical linear models. Mixed models can account for ordinal data as well as continuous data.

Because our study design violated the assumption of independence, mixed models were performed to investigate the relationship between mental workload and other variables of interest (including questionnaire data and phone usage data). Because of the input requirement for detailing the predictor variable and outcome variable, mental workload levels were assigned as the predictor variable, due to our primary aim of investigating the impact that mental workload has on our lives. One exception was sleep data, which was investigated as both the outcome and predictor variable, in order to investigate not only the effect of mental workload on sleep the following night, but also the effect that sleep quality and duration the night before may have on people's mental workload abilities the following day.

In order to run the models, each pair of variables were first investigated by fitting a regression model to determine whether a linear, quadratic, or cubic fit produced the least error in the model. The residuals were then evaluated to determine the distribution of the data. Our data contained Gaussian distributions and Beta distributions. For Gaussian distributions, linear mixed models (LMMs) were performed. For Beta distributions, generalised mixed models (GLMMs) were performed, in which the algorithm required the data to be standardised between 0-1. Due to the non-independence of the data, participant was inputted as a random effect; participant and hour were inputted as random effects for the analysis of phone usage data only. Standardized parameters were obtained by fitting the model on a standardized version of the dataset. The model was estimated using REML and nloptwrap optimizer for Gaussian distributions and ML and nlminb optimizer for Beta distributions. 95% Confidence Intervals and p-values were computed using lme4 and glmmTMB for LMMs, and the Wald approximation for the GLMM output.

The output of the models included the R squared values, the beta values, 95% confidence intervals, and the P value of the model.

For significant findings, scatterplots with a simple regression line were produced in order to visually display the data trends. However, these plots display one regression line, whereas the mixed models apply a different line for each random effect (participant), which means the visualisations do not capture the full and accurate story.

Participant 15 was excluded from the analysis of the questionnaire data due to poor data response.

5.4 Quantitative Findings

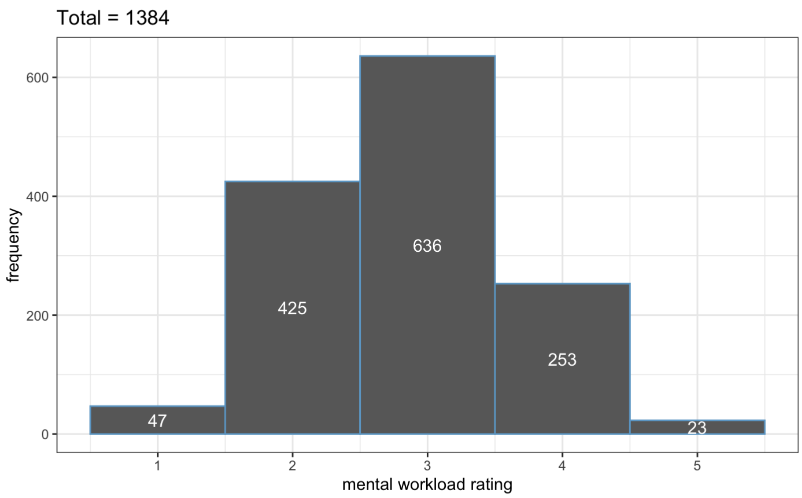


Figure 5.2: Frequency of each mental workload rating.

Figure 5.2 outlines the frequency of the different mental workload ratings from all participants across the whole 5 days of data collection. As the graph depicts, out of a total of 1384 ratings mental workload, participants rated level 3 the most frequently with 636 ratings, level two was next with 425 ratings, level 4 then followed with 253 ratings, level 1 was next with 47 ratings, and then was level 5 with 23 ratings.

It is notable how few ratings were made for levels 1 and 5, which were the extreme ends of the rating scale. When exploring the use of the rating scale, only 3 participants used the full rating 1-5 rating scale during their week of data collection; 10 participants used a rating of 1, and 7 participants used a rating of 5. 18 participants used ratings 2, 3, and 4, and 1 participant did not input a rating of 4.

Diary data gave an indication of the types of activities that participants inputted alongside their mental workload level ratings. These are outlined in word clouds (Figures 5.3a, 5.3b, 5.4a, 5.4b, and 5.4c).

Because the qualitative interviews were discussed in terms of 'low', 'medium', and 'high' mental workload, the data presented in this chapter considers ratings of 2 as 'low' mental workload, ratings of 3 as 'medium' mental workload, and ratings of 4 as 'high' mental workload in order to enable comparisons between the qualitative and quantitative finding. Because of the terminology of the ISA scale [36], ratings 1 and 5 were interpreted as underload and overload respectively, despite not specifically stating this in the scale terminology; the ratings in between seemed to align smoothly with the levels in between these extremes.

(a) Rating 1.

(b) Rating 2.

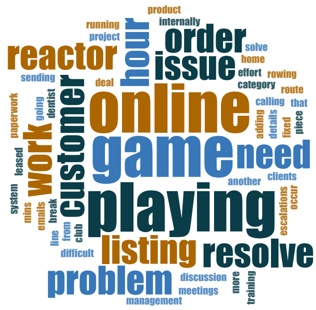
Figure 5.3: Word clouds of diary data associated with mental workload ratings 1 and 2. Words included were a minimum of 4 letters in order to largely filter out contaminating filler words such as 'the.'

5.4.1 Patterns of Fluctuations

Figure 5.5 broadly summarises the ratings data in a boxplot that outlines all the weighted average ratings for each hour, across all participants and

(a) Rating 3.

(b) Rating 4.



(c) Rating 5.

Figure 5.4: Word clouds of diary data associated with mental workload ratings 3, 4, and 5.

days. The most concentrated rating hours were between 7am and 10pm. Within this, there was a large distribution of ratings between the different mental workload levels. Generally, the plot depicts that average mental workload levels were lower before work, highest at work, peaking between 10am and 12pm, and then dropped lower again after work; average levels did not reach much higher than a level of 3.

The qualitative findings identified participants' preferred patterns of fluctuations between the mental workload levels (*The Mental Workload Cycle* subtheme). We therefore explored the types of transitions that were present in their subjective ratings data over the 5 days of data collection. Figure 5.6 displays the proportion of transitions for all subjective data, i.e. the rating after the previous rating.

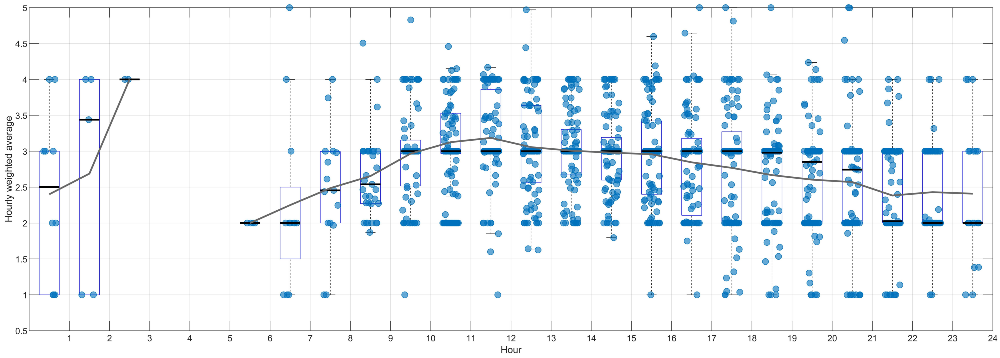


Figure 5.5: Boxplot of the weight average mental workload ratings for each hour. The continuous line represents the mean. The values on the x-axis represent the hour of the day, e.g. the data points between 11-12 represents data between 11am-12pm. As depicted, there were no ratings between 3am-5am on any of the days.

As shown in the diagram, the most frequent transition was the rating of another 3 following a 3, supporting the sustainability of medium mental workload. The data reflects lots of transitions between ratings 2, 3, and 4, which would be expected based on the qualitative findings. In terms of the patterns of fluctuations between these levels, participants transitioned more frequently from low mental workload levels to medium levels compared to high levels, though both transitions were often represented in the data. From a medium level, participants transitioned to both low and high mental workload levels frequently, reflecting the individual qualitative preferences that participants had about whether they would seek up or down after a medium level. From high levels, participants transitioned more frequently to medium levels than to low levels; this is the opposite of what the qualitative findings suggested.

From Medium to:

The qualitative findings identified that participants found medium mental workload to be sustainable, but still had their own preferences about which

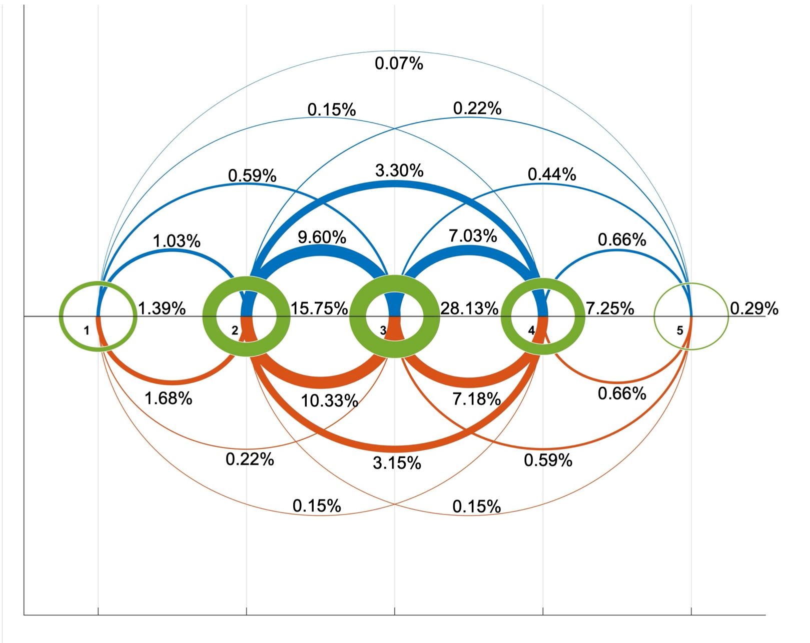


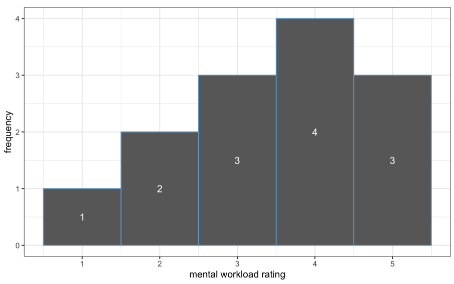
Figure 5.6: Diagram shows the amount of transitions that participants made between mental workload levels with their corresponding percentage. The upper half of the diagram (blue) represents 'upwards' transitions (such as transitioning from a 1 to a 2) and the bottom half (red) represents 'downward' transitions (such as from a 2 to a 1). The circles (green) around the rating levels represent the transition to the same level (such as a 1 to another 1). The thickness of the lines represent the frequency of the transition.

mental workload level they would transition to after time spent at a medium mental workload level (*The Mental Workload Cycle* subtheme). This allowed for the comparison between the stated qualitative preferences and the quantitative patterns in the mental workload ratings data. From the qualitative findings, 2 participants did not explicitly state their preference, and thus the comparisons include the data from the 17 participants who stated their transition preference. The visualisations that are presented outline the frequency of ratings that occurred following a 3 (a medium mental workload level); unsurprisingly, there were many instances of a 3 rating following a 3, and then the visualisations reflect the frequency of the other ratings.

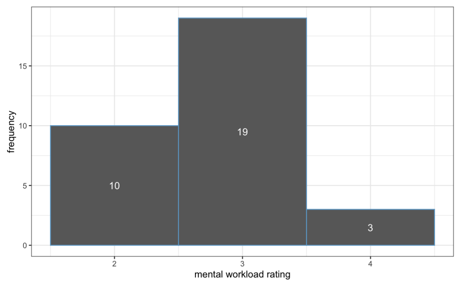
11 participants expressed that they would seek to spend time at a higher

mental workload level after a medium. Out of these, only 1 participant's data reflected that (Figure 5.7a). 6 participants' data actually showed more low ratings after medium ratings (Figure 5.7b), and 4 participants had an even split of ratings between low and high (with a maximum difference of 2 ratings between high and low levels).

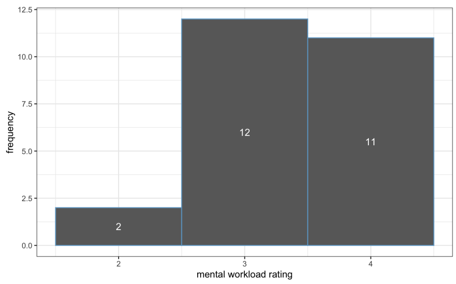
6 participants outlined that they would prefer to spend time at lower mental workload levels after medium mental workload. Out of these, no participants achieved that. 3 participants clearly rated more high mental workload after medium ratings (Figure 5.7c), and 3 participants had even ratings between high and low levels (with a maximum difference of 2 ratings between them).



(a) P12 had a preference for transitioning to high mental workload after a medium mental workload, which is reflected in the data.



(b) P4 had a preference for transitioning to a high mental workload after a medium mental workload, but their data reflects more transitions to low mental workload.



(c) P19 preferred to transition to a low mental workload after a medium mental workload, but their data reflects more transitions to high mental workload.

Figure 5.7: Transition patterns after medium mental workload level.

Summary

This section considered mental workload ratings of 2, 3, and 4 as mental workload levels corresponding to low, medium, and high. There were noticeable observable similarities and differences between the activities that were reported for each mental workload level. Participants generally had the highest mental workload levels in their days during work hours from around 9am-4pm. There were lots of transitions between low, medium, and high mental workload levels, which supported the qualitative findings of frequent fluctuations; the most frequent transition between ratings was from a medium level to another medium level, which supported the qualitative findings regarding the sustainability of medium mental workload. However, the patterns of fluctuations were often not in the preferred patterns that participants had qualitatively expressed. Notably, from a medium mental workload level, participants frequently appeared to to transition in the opposite direction to their qualitative preference. Reasons and implications for these discrepancies will be discussed further in Section 5.5.

5.4.2 Questionnaire Analysis

A weighted average was calculated for each participant for each day. The values ranged between 1.84 and 3.85 (mean = 2.87, standard deviation (SD) = 0.43). Ratings are displayed in Figure 5.8.

The qualitative findings identified the negative consequences that could arise from sustaining any mental workload level for too long (*Sustainment is an Issue*). To further the exploration into the impact that mental workload has on our lives, we ran mixed models to investigate the how different overall amounts of mental workload effect the factors measured by the

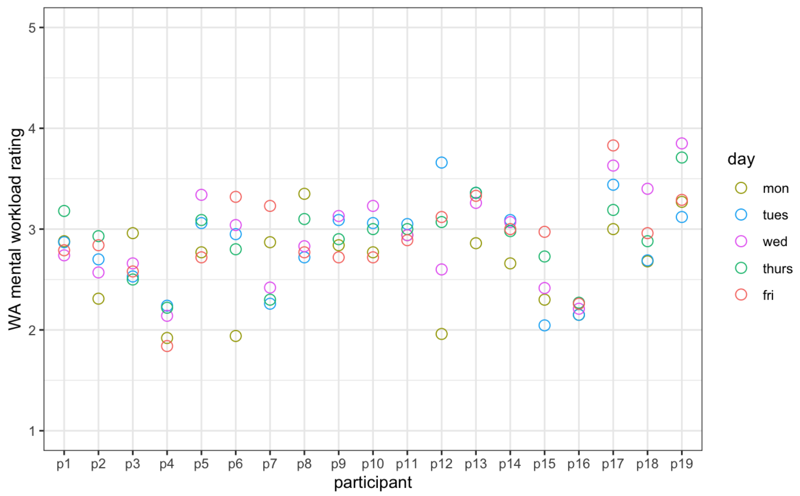
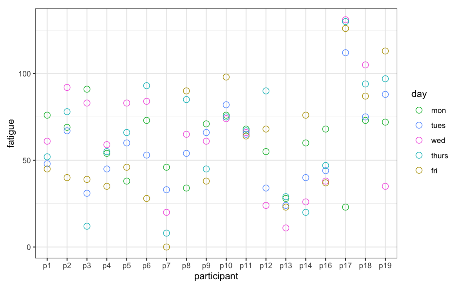


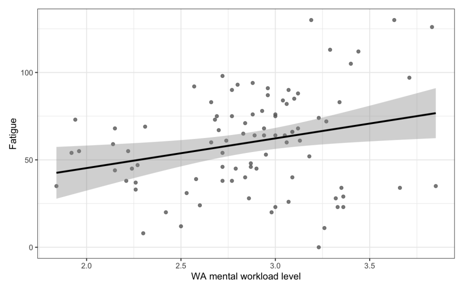
Figure 5.8: Weighted average mental workload ratings for each participant for each day.

evening questionnaires. The data for each questionnaire is outlined in turn.

Fatigue



(a) Fatigue ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.



(b) Mental workload levels could significantly estimate perceived fatigue levels at the end of the day. The relationship was linear, as shown by the line.

The scores for the fatigue section of the VAS could be between 0 and 143. Participants responses ranged between 0-130, mean = 60.01, SD = 28.17 (Figure 5.9a).

A linear model was the best fit for the data. We fitted a linear mixed model with mental workload as the predictor and fatigue as the outcome variable. (Figure 5.9b). The model included participant as random effect.

The model's total explanatory power was substantial (conditional R2 = 0.36) and the part related to the fixed effects alone (marginal R2) was of 0.05. Within this model:

The effect of mental workload was statistically significant and positive (beta = 13.89, 95% CI [-0.02, 27.81], t(86) = 1.98, p = 0.050; Std. beta = 0.21, 95% CI [-3.29e-04, 0.43]).

To add explanation to these values, the fixed effects R2 value (0.05) indicates the amount of variance (5%) in the outcome variable (fatigue) caused by the predictor variable (mental workload). The beta value (13.89) is the slope coefficient and indicates the amount of change in the outcome variable for every one unit of change in the predictor variable. The 95% confidence intervals next to this defines the range of values that the slope coefficient plausibly falls into. The t-value (86) and its significance value (p = 0.050) then concerns whether the slope coefficient value is significantly different to 0. After this, the standardised beta (0.21) and its confidence intervals enable the comparison between predictors if they vary in measurement units; standardised betas show standard deviations, such that for 1 standard deviation change in mental workload level, there is an expected 0.21 standard deviation increase in fatigue.

Thus, these results suggest that as daily mental workload levels increased, fatigue at the end of the day also increased. This is perhaps an intuitive finding, but appears to have limited previous research.

Energy

The scores for the energy section of the VAS could be between 0 and 55. Participants responses ranged between 0-41, mean = 18.17, SD = 8.88

(Figure 5.10).

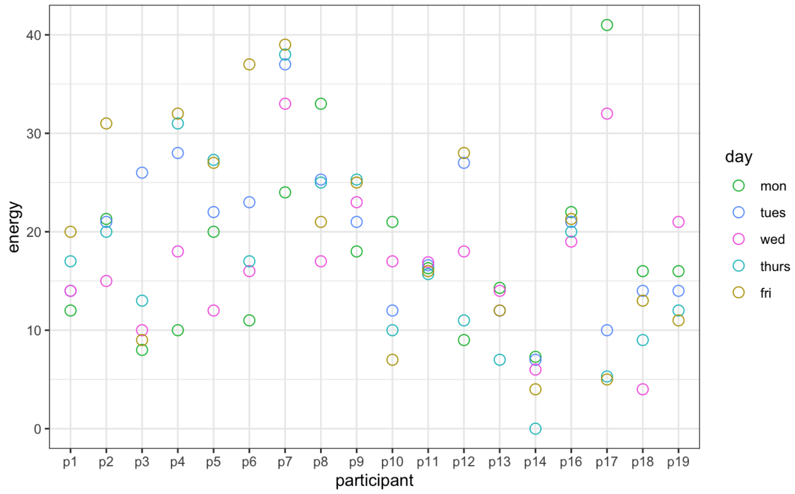


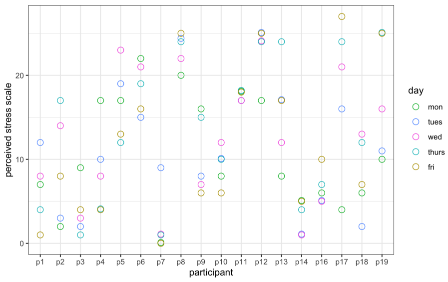
Figure 5.10: Energy ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.

We ran the linear mixed model with a linear fit for energy and mental workload and the effect of mental workload was statistically non-significant and positive (beta = 0.06, 95% CI [-4.35, 4.46], t(86) = 0.03, p = 0.979; Std. beta = 2.80e-03, 95% CI [-0.21, 0.22]). Thus, our quantitative findings did not find a significant relationship between daily mental workload levels and energy levels at the end of the day. This is perhaps surprising as it could be expected that as mental workload was associated with increased fatigue, it would also be associated with decreased energy levels. Indeed, a Spearman's correlation found a significant negative correlation between fatigue and energy, rs = -.385, p = .001.

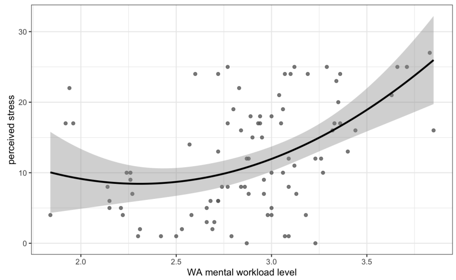
Perceived Stress Scale

The response to the PSS could be between 0-40. The responses from participants ranged between 0-27, mean = 12.09, SD = 7.80 (Figure 5.11a).

For stress, a quadratic model was the best fit for the data. We fitted a



(a) PSS data for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.



(b) Mental workload levels could significantly estimate perceived stress levels at the end of the day. The relationship was quadratic, as shown by the line.

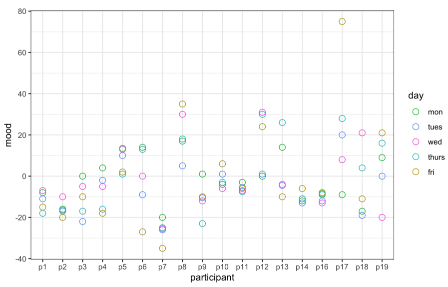
linear mixed model with mental workload as the predictor variable and PSS as the outcome variable (Figure 5.11b). The model included participant as random effect. The model's total explanatory power was substantial (conditional R2 = 0.65) and the part related to the fixed effects alone (marginal R2) was of 0.11.

The effect of mental workload was statistically significant and positive (beta = 1.06, 95% CI [0.53, 1.59], t(86) = 3.97, p = .001; Std. beta = 0.33, 95% CI [0.17, 0.50]).

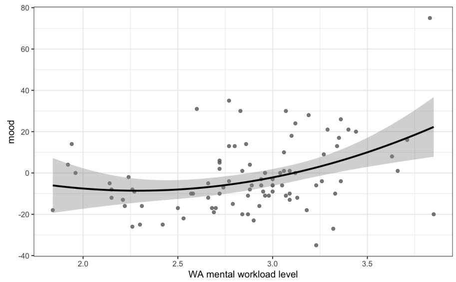
Thus, mental workload explained 11% of the variance in the perceived stress data. This is interesting, as it suggests that increased stress levels are associated with too much low mental workload as well as too much high mental workload. Speculatively, reasons for this might include people find that avoiding work causes increase stress, or that not finding work or daily tasks mentally challenging is a source of stress to people.

Mood Disturbance

POMS responses could be between -44 and 116 (higher scores indicate more negative mood). Participants responses ranged between -35 and 75, mean



(a) Mood disturbance ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.



(b) Mental workload levels could significantly estimate perceived mood disturbance (higher values correspond to a more negative mood) at the end of the day. The relationship was quadratic, as shown by the line.

= 2.01, SD = 17.16 (Figure 5.12a).

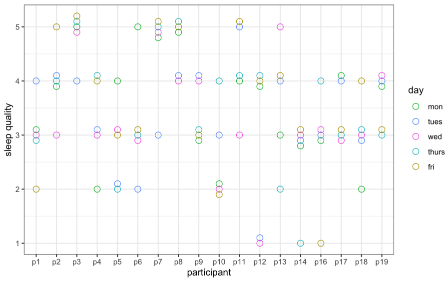
We fitted another linear mixed model with a quadratic fit, where mental workload was the predictor and mood disturbance was the outcome variable. (Figure 5.12b). The model included participant as random effect. The model's total explanatory power was substantial (conditional R2 = 0.40) and the part related to the fixed effects alone (marginal R2) is of 0.06. Within this model:

The effect of mental workload was statistically significant and positive (beta = 1.62, 95% CI [0.19, 3.05], t(86) = 2.25, p = 0.027; Std. beta = 0.23, 95% CI [0.03, 0.44]).

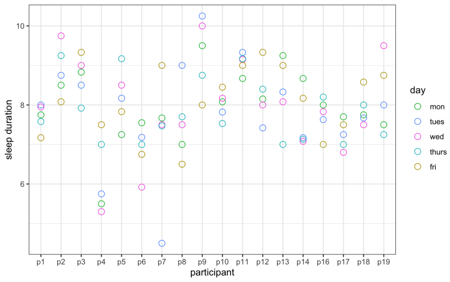
To a significant extent, mental workload levels during the day were associated with mood disturbance in the evening. Similarly to the stress findings, more negative moods were associated with both low and high daily levels of mental workload. This could also speculated to be for similar reasons as the stress findings, as these factors seem related in terms of negative perceptions. In this respect, a Spearman's correlation found a significant positive correlation between stress and mood disturbance, rs = .750, p = 0.001.

Sleep Quality, Duration, and Onset - Following Night

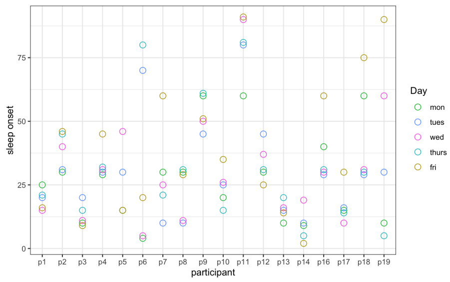
Sleep quality scores could be between 1-5. Participants responses ranged between 1-5, mean = 3.46, SD = 1.05 (Figure 5.13a). Sleep duration responses ranged between 4.5 hours to 10.25 hours, mean = 7.94, SD = 1.02 (Figure 5.13b). Sleep onset time ranged between 2-91 minutes, mean = 31.87 minutes, SD = 21.86



(a) Sleep quality ratings for each participant for each day.



(b) Sleep duration data for each participant for each day.



(c) Time to fall asleep after settling down data for each participant for each day.

Figure 5.13: Sleep data for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.

Assessing the impact of daily mental workload levels on sleep in terms of quality and onset was the area of investigation that had been most robustly grounded in previous research.

The linear mixed model was ran to investigate the impact of mental workload on sleep duration. A quadratic fit was most suitable for the model, but the effect of mental workload on sleep duration was statistically non-significant and positive (beta = 0.03, 95% CI [-0.05, 0.12], t(86) = 0.74, p = 0.464; Std. beta = 0.08, 95% CI [-0.13, 0.28]).

As this was real world data, it is unsurprising that this result was non-significant. It can be assumed that participants were not left to sleep for as long as they naturally would because of their morning commitments to e.g. childcare or work.

The linear mixed model was performed to investigate the impact of mental workload and sleep quality (the following night). The best fit for the model was linear. The effect of mental workload was statistically non-significant and negative (beta = -0.13, 95% CI [-0.67, 0.41], t(86) = -0.49, p = 0.623; Std. beta = -0.06, 95% CI [-0.28, 0.17]).

A generalised mixed model was ran with a linear fit for time to all asleep after settling down and mental workload. The effect of mental workload was also statistically non-significant and positive (beta = 0.29, 95% CI [-0.34, 0.92], p = 0.364; Std. beta = 0.13, 95% CI [-0.15, 0.40]).

These latter findings are not what would have perhaps been expected as they suggest no relationship between mental workload and sleep quality or onset. The discussion section interprets these results further based on previous research.

Sleep Quality and Duration - Previous Night

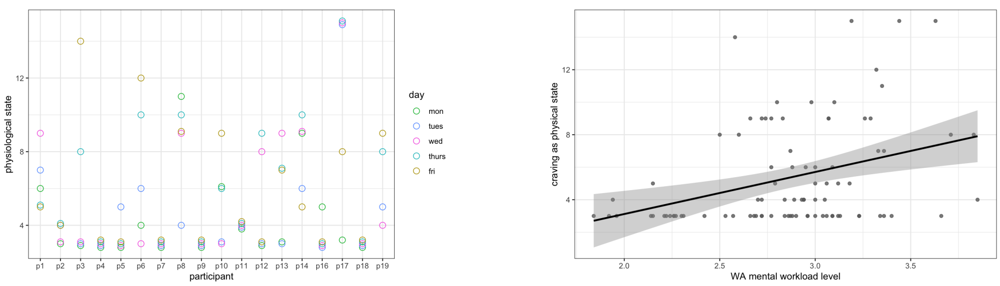
A linear mixed model with a linear fit was performed to investigate the effect of sleep quality and duration on mental workload the following day. The effect of sleep duration was statistically non-significant and negative (beta = -9.18e-03, 95% CI [-0.29, 0.28], t(66) = -0.06, p = 0.949; Std. beta = 0.18, 95% CI [-0.04, 0.40]). The effect of sleep quality was also statistically non-significant and negative (beta = -0.18, 95% CI [-0.84, 0.48], t(66) = -0.55, p = 0.585; Std. beta = 0.01, 95% CI [-0.17, 0.20]).

In line with the findings relating to the effect of mental workload of sleep the following night, when sleep was used as the predictor variable to investigate the effect of sleep the previous night on mental workload the following day, no significant relationship was found.

Food Cravings

Each subsection of cravings data responses could be between 3-15.

Craving as a Physiological State Responses ranged between 3-15, mean = 5.35, SD = 3.21 (Figure 5.14a).



(a) Craving as a physiological state ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.

(b) Mental workload levels could significantly estimate physical craving levels at the end of the day. The relationship was linear, as shown by the line.

We fitted a generalised mixed model to investigate the effect of mental workload on cravings as a physiological state. The best fit was a linear model.

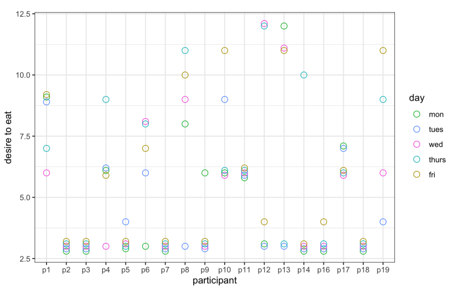
(Figure 5.14b). The model included participant as random effect. The model's total explanatory power was substantial (conditional R2 = 0.29) and the part related to the fixed effects alone (marginal R2) is of 0.11.

Within this model:

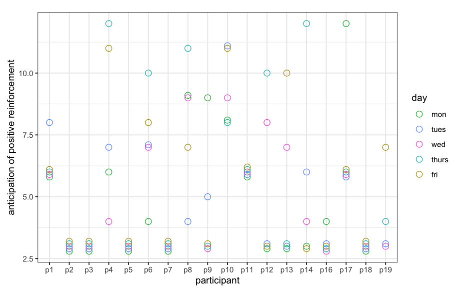
The effect of mental workload was statistically significant and positive (beta = 0.86, 95% CI [0.22, 1.50], p = 0.008; Std. beta = 0.37, 95% CI [0.10, 0.65]).

Interestingly, this finding indicated that mental workload could explain 11% of the variance in the physiological cravings data. The three questions related to this type of craving were: ‘I am hungry,’ ‘If I ate right now, my stomach wouldn’t feel as empty,’ and ‘I feel weak because of not eating.’ This can reasonably interpreted as feelings of hunger, such that daily mental workload levels were linearly associated with hunger levels at the end of the day. Participants were instructed to fill out the questionnaires at the same time each day after their evening meal, meaning the results should not be because of differing situational factors. Thus, it could be speculated that like physical activity that burns energy and hence requires the body to refuel properly, there may be a similar effect (perhaps a perceived effect) with mental workload.

An Intense Desire to Eat Responses ranged between 3-12, mean = 5.38, SD = 2.85 (Figure 5.15a).



(a) Intense Desire to Eat ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.



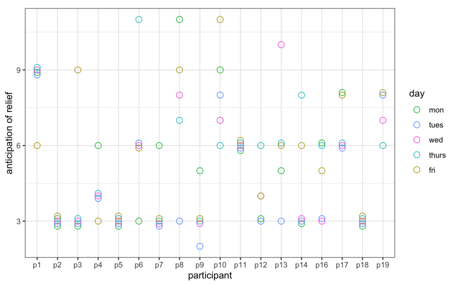
(b) Relief from eating ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.

A linear mixed model that accounted for the beta distribution of the data was performed with a linear fit for an intense desire to eat and mental workload. The effect of mental workload was statistically non-significant and positive (beta = 0.58, 95% CI [-0.07, 1.23], p = 0.083; Std. beta = 0.25, 95% CI [-0.03, 0.53]).

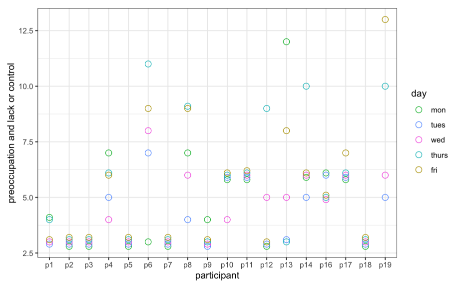
Relief from Negative States and Feelings as a Result of Eating Results ranged between 3-12, mean = 5.31, SD = 2.78 (Figure 5.15b).

The same was performed for anticipation of relief from negative states and feelings as a result of eating and mental workload. The effect of mental workload was statistically non-significant and positive (beta = 0.52, 95% CI [-0.07, 1.11], p = 0.081; Std. beta = 0.23, 95% CI [-0.03, 0.48]).

Obsessive Preoccupation with Food or Lack of Control Over Eating Responses ranged between 3-11, mean = 5.20, SD = 2.38 (Figure 5.16a).



(a) Food preoccupation ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.



(b) Anticipation of reinforcement ratings for each participant for each day. Overlapping data points of the same values have been shifted slightly for visualisation purposes.

The same model was applied for obsessive preoccupation with food or lack of control over eating and mental workload. The effect of mental workload was statistically non-significant and positive (beta = 0.38, 95% CI [-0.27, 1.03], p = 0.254; Std. beta = 0.16, 95% CI [-0.12, 0.45]).

Anticipation of Positive Reinforcement that may Result from Eating Responses ranged between 3-13, mean = 5.01, SD = 2.32 (Figure 5.16b).

The same model but with a cubic fit was performed for anticipation of positive reinforcement that may result from eating and mental workload. The effect of mental workload was statistically non-significant and positive (beta = 0.01, 95% CI [-0.02, 0.04], p = 0.449; Std. beta = 0.11, 95% CI [-0.18, 0.40]).

Thus, the rest of the cravings data showed no significant relationship regarding the impact that mental workload had on aspects of cravings.

Summary

The results in this section outlined the relationship between mental workload and the questionnaire data. There were four significant findings. Daily mental workload levels were found to contribute to fatigue, stress, mood disturbance, and feelings of hunger at the end of the day. For fatigue and hunger, this relationship was positive an linear, such that increasing daily mental workload levels were associated with increasing fatigue and hunger levels. The relationship was quadratic for stress and mood disturbance, where low and high daily mental workload levels were associated with higher stress and more negative mood when measured during the evening.

5.4.3 Mobile Phone Usage

The qualitative findings identified that periods of high mental workload were associated with feeling less distracted, and periods of low mental workload were associated with feeling more distracted (*General Perceptions of Mental Workload* theme). In relation to this, we analysed phone usage data as a indication of distractedness.

The two categories of interest were time spent on social media, and total phone usage time (minus system usage data, which included background apps etc). The phone usage data was available to analyse in one hour chunks. Social media usage time ranged between 1 second an hour and 2598 seconds (43.30 minutes) an hour, mean = 315.69 seconds (5.26 minutes), SD = 362.82. Total usage time ranged between 63 seconds (1.30 minutes) an hour and 2613 seconds (43.55 minutes) an hour, mean = 905.61 seconds (15.09 minutes), SD = 719.67.

Initial inspection of the data indicated a relationship between mental workload ratings and phone usage. Figure 5.17 shows the initial visualisation that was plotted for P1 on their first day of data collection. The plot shows that when mental workload levels rise, phone usage time drops, and when mental workload decreases, phone usage increases.

We then statistically evaluated the relationship across all participants between mental workload ratings and social media usage in terms of the effect that weighted average mental workload levels at each hour of social media usage had on social media usage. A mixed model was performed to assess this relationship with hour of day added as a random effect alongside participant. (Figure 5.18a).

A generalised mixed model was fitted due to the beta distribution of the

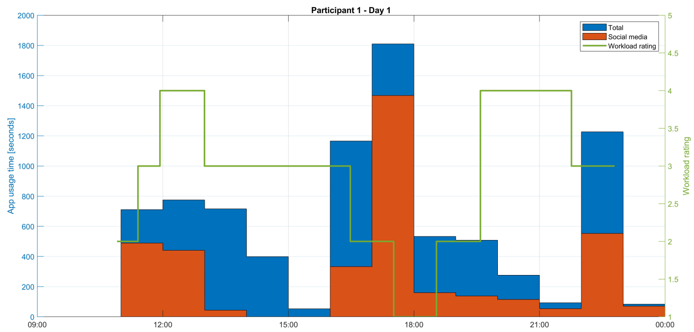
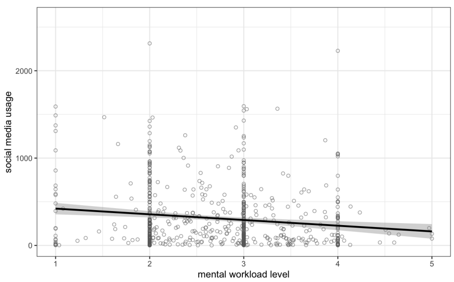


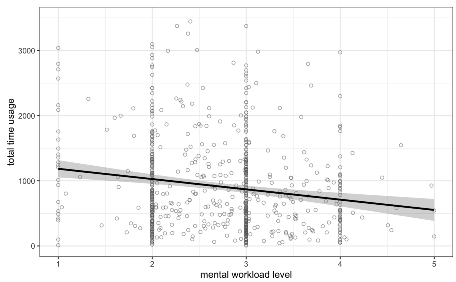
Figure 5.17: A plot of mental workload ratings and phone usage time for P1, day 1. A general trend can be seen as mental workload ratings become lower, phone usage time becomes higher, and vice verse.

data. A linear fit was most suitable for the model. Mental workload was again used as the predictor variable and social media usage was used as the outcome variable. The model's total explanatory power was moderate (conditional R2 = 0.20) and the part related to the fixed effects alone (marginal R2) was of 0.02. Within this model:

The effect of mental workload was statistically significant and negative (beta = -0.15, 95% CI [-0.25, -0.05], p = 0.004; Std. beta = -0.11, 95% CI [-0.19, -0.04]).



(a) Mental workload levels each hour was significantly related to social media usage.



(b) Mental workload levels each hour was significantly related to total phone usage time.

The relationship between mental workload and total phone usage was investigated in the same way (Figure 5.18b). The model's total explanatory power was substantial (conditional R2 = 0.33) and the part related to the fixed effects alone (marginal R2) was of 0.02. Within this model:

The effect of mental workload was statistically significant and negative (beta = -0.13, 95% CI [-0.23, -0.03], p = 0.010; Std. beta = -0.10, 95% CI [-0.17, -0.02]).

Although there is a small amount of variance in the data explained by hourly mental workload levels for both social media usage and total phone usage, these findings were significant and indicate that there is a reliable relationship that shows mental workload levels do contribute to phone usage. If this is considered in terms of distractions, it can be speculated that mental workload would be a large predictor of time spent doing distracting activities if they were all added into the model. This trail of speculation is expanded in the discussion section.

Because some data for social media time usage indicated that for some hours, participants had spent 1 second on social media, this model was also performed for data starting at 10 seconds of social media usage. The output was the same as the first model. As a following exploration, we wanted to evaluate the relationship between very short uses of social media (up to 2 or 5 seconds) on mental workload levels. The quantity of data we had was not sufficient for statistical analysis, and a boxplot did not indicate much difference between short uses and longer uses (Figure 5.19).

5.5 Discussion

This chapter investigated mental workload in daily life with the aim of furthering our understanding of how tracking such "brain data" could be used to improve our work performance and lives. Three sections contributed to this. Firstly, the qualitative analysis of the contextualisation section provided insight into the overarching reasons for why participants believed that

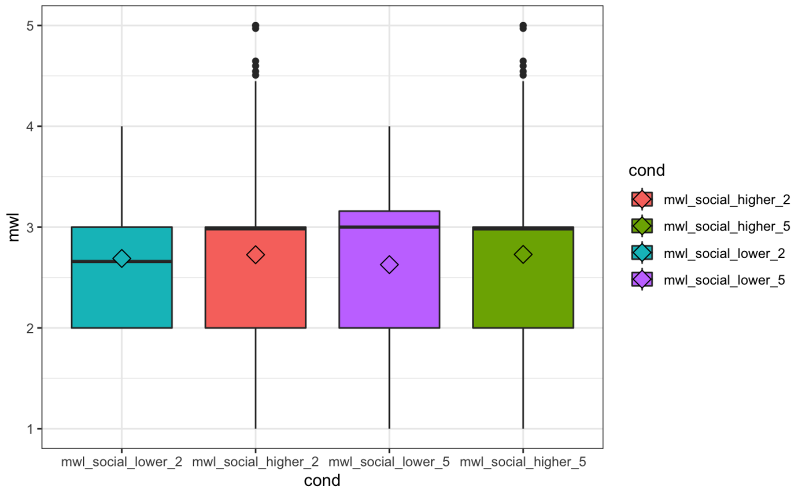


Figure 5.19: Boxplot showing descriptives for different social media usage time conditions on mental workload ratings, including usage for 2 seconds or less and 3 seconds or more, and usage for 5 seconds of less and 6 seconds or more. Mean mental workload rating values are indicated by the position of the diamonds.

tracking this type of data might be valuable to them as a form of personal informatics. This section was brief and broad, provided validation for interest in tracking cognitive activity, and perhaps set the scene for a deeper insight into mental workload provided by the following two sections. The qualitative analysis of the personal experiences section investigated lived experiences of mental workload, and the quantitative analysis investigated the function and impact of mental workload from a longitudinal life perspective; these sections provided a rich insight into mental workload as a concept.

We aimed to build upon the qualitative analysis by designing the quantitative analysis around the findings from personal experiences section, which had provided an initial understanding of the function of mental workload in daily life and how it may be useful as a form of personal informatics.

More specifically, four themes from the personal experiences section were identified in regards to this. Theme 1 outlined that there were different general perceptions of high and low mental workload levels, in that different participants had either positive or negative perceptions of the same level. In contrast, all participants regarded medium mental workload positively. Indeed, the quantitative data in terms of the mental workload ratings app showed that medium mental workload (rating 3) was rated most frequently, arguably supporting medium mental workload as a positively regarded revel.

Theme 2 identified four factors (pressure, enjoyment, outcome, and location) that could change the initial perceptions of the high and low mental workload levels from positive to negative, or vice verse. These themes further our understanding of how experiences of mental workload are perceived.

Theme 3 presents an apparent Mental Workload Cycle, where findings identified the necessity of fluctuating between mental workload levels which tended to occur in specific patterns. Fluctuations prevented sustaining any level for too long as this could have negative consequences on oneself or work output (e.g. fatigue or feeling unproductive); medium mental workload was considered the most sustainable level, but still appeared to result in negative consequences after a longer period of time. Fluctuating between levels also allowed individuals to benefit from the different positive characteristics of each level (e.g. a sense of achievement or time to recover). The findings from this theme contribute largely towards understanding the impact that mental workload may have on our lives and work (and the impact that our lives and work may have on our mental workload performances), and increases our understanding of what we should aim for in terms of mental workload. Theme 4 identified three factors (life, internal and external) that could prevent individuals from achieving their Cycle fluctuations by decreasing the opportunity to fluctuate or affecting the ability to operate at certain levels.

The quantitative data was analysed predominantly in relation to the Mental Workload Cycle. Patterns of fluctuations showed frequent transitions between low, medium, and high levels. The data supported medium mental workload as the most sustainable level by showing that transitioning from a medium level (rating 3) to another medium level was the most frequent transition in the data set (28.13%). The patterns of transitions, however, did not always align with the qualitative findings in terms of participant preferences. Additionally, the questionnaire data supported the findings regarding the negative implications that can occur from spending too much time at certain mental workload levels, and phone usage data could somewhat be useful for indicating mental workload levels based on level of distractedness. These findings are discussed more below.

Whilst our main contributions derive from qualitative themes 3 and 4, themes 1 and 2 are important for two reasons. Firstly, as far as we are aware, no research has investigated mental workload from a ‘people perspective,’ in terms of how mental workload is qualitatively conceptualised by those who experience it; there is a large body of mental workload research [204; 240; 176], but the focus remains on isolated tasks measuring quantitative data. Increasing our understanding of the experiences behind the numbers may contribute to greater progress in these research areas, such as increased understanding of what contributes to overload and underload. Secondly, themes 1 and 2 lay the foundations for understanding people’s approaches to mental workload in their lives, such as high mental workload avoidance because of associations with stress, or which level people prefer to transition to after a medium mental workload level. When developing the Mental Workload Cycle, having an understanding of the different perceptions of mental workload enabled a richer insight into why having mental workload fluctuations are preferable, how Cycle preferences may vary between individuals, and how preferences do not always translate to observed behaviour, discussed in more detail below.

5.5.1 Personal informatics and BCIs

Current pBCI neurotechnology available to consumers (to help people e.g. focus or meditate) are tailored around helping users to achieve a certain state in the present moment. What we have investigated is how tracking mental workload data over longer periods of time (days/weeks/months) could contribute towards making improvements in our lives, as a form of personal informatics [82; 145; 194; 196].

Mental workload was chosen as a concept that is fundamental in our daily lives and has a large body of research aimed at accurately tracking it in the real-world. It is likely that this tracking technology will be available as type of pBCI in the relatively near future [14; 16]. Its current application applies to improving performance at work, especially in safety-critical jobs [16; 240], but our qualitative findings suggest that tracking mental workload from a broader life perspective could have positive implications for our wellbeing and performance on tasks. This is because if we keep track of our mental workload levels and aim to adhere to the Mental Workload Cycle that is optimal for us in terms of fluctuations between levels, we could avoid the negative consequences that come from sustaining levels for too long, and reap the rewards of the benefits that each level can have on our work and lives.

However, as seen in the quantitative analysis, optimal fluctuation patterns, as well as preferred fluctuation patterns, were not always achieved. Firstly, participants expressed that they would ideally transition from a high mental workload to a low mental workload. Whilst this did occur in the data, participants more often transitioned to a medium mental workload from a high mental workload. Additionally, participants had their own preferences about which level they would ideally transition to after a medium mental workload level. However, the data indicated that only one participant's actual behaviour followed their expressed preference; the ratings data often indicated that participants transitions from a medium level were the opposite of their preference, or at least evenly distributed between low and high transitions.

These findings suggest a dissociation between participant preferences and actual behaviour. Speculatively, this could be explained in a few ways. Firstly, because the ratings app data represents 'overall' mental workload levels since the last rating, it is sensible to consider that fluctuations within that time period often occurred and were lost within the rating, meaning actual transitions may not have been represented fully in the data. However, on a broader level, the data does provide insights into mental workload fluctuations. This is especially clear in the data that shows the participants who preferred to go to a high mental workload level after a medium level frequently went to a low level instead, whereas the participants that preferred to go to a low level frequently went to a high level instead.

It is therefore also conceivable that participants qualitatively expressed behavioural intentions that they believed were desirable for them to do, but did not actually align with their behaviour in practice. This could be attributed to a general discrepancy between desirable intentions and actual behaviour, or participants being unable to facilitate their preferences in practice. In relation to a general discrepancy between desirable intentions and actual behaviour, there are many instances of inconsistencies between intentions and actions in a wide variety of domains, referred to as the intention-behaviour gap [205] or a hypothetical bias [6]; an example of this is in the health domain, where people often intend to participate in an exercise or rehabilitation programme, but not act on this intention. This general discrepancy is believed to be explained by the belief-disparity hypothesis [6; 7]. This outlines how hypothetical and real-life contexts are construed in very different ways, where favourable beliefs and attitudes are expected in hypothetical contexts because of the lack of salient contextual cues. In this sense, it could be hypothesised that participants expressed their transition preferences based on what they believed was the most desirable behaviour as opposed to their commonly performed actions. For example, believing it is more productive to reach a high mental workload after a medium but actually tending to drop down to a low level; or believing it is healthier to drop to a low level after a medium level, but commonly actually operating at a higher level.

Alternatively, participants may not have been able to facilitate their intentions in practice because of their circumstances. Theme 4 of the qualitative findings outlined how there are internal, external, and life factors that can prevent people from operating at their intended or desired mental workload levels. In this case, the discrepancy between the preferred (qualitative) and observed (quantitative) findings could be because of situational contexts that did not enable participants to execute their preferred fluctuations or levels, such as the demands of their jobs. In relation to this, previous research found that even though participants were educated about how many hours sleep they should get each night, many did not translate this into action because of their circumstances [106]. The authors suggested that help to implement strategies and modify lifestyle factors was important for helping people to achieve their sleep goals. Research like this alongside our qualitative findings in theme 4 suggests that even if we know what we 'should' typically aim for in terms of mental workload, there may need to be other goals set to prevent the negative outcomes that may arise from not fluctuating between levels in those people who struggle to achieve the ideal fluctuations because of their circumstances.

The two potential reasons outlined above regarding why actual behaviour may not have aligned with preferred behaviour indicate that being able to reflect on mental workload data and be guided by actionable insights could help people to achieve their mental workload aims in terms of facilitating their Mental Workload Cycle. As mentioned, our qualitative findings suggest that adhering to a Cycle could prevent the negative consequences that arise from sustaining any level for too long. In this sense, perhaps people would, for example, feel less burnt out and resentful, and feel more rested and efficient. A large part of the quantitative analysis also regarded investigating the impact that overall mental workload levels during a day had on participants' daily lives. The weighted average value that represented the 'amount' of mental workload that participants had experienced in their day, enabled an understanding of how experiencing a lot of, e.g., high or low mental workload during the day may impact certain wellbeing or cognitive aspects. The qualitative analysis identified several factors that experiencing too much of any mental workload level could negatively impact; again, medium mental workload was considered the most sustainable, with participants generally content to experience a lot of that level in their day, despite indications that too much medium mental workload could also lead to negative consequences. The quantitative phase was designed before the findings of the qualitative phase, meaning that the measures could not be guided in this regard. However, there were some measures taken during the quantitative phase that emerged as findings in the qualitative phase; three of the four significant findings in the quantitative analysis were identified qualitatively as negative effects from sustaining different mental workload levels for too long. These included stress, fatigue, and mood.

In terms of stress, the qualitative findings identified that too much high mental workload was specifically associated with increased stress; too much low mental workload was associated with decreased enjoyment, decreased productivity, and decreased satisfaction, which are factors that could arguably be associated with feelings of stress. The quantitative findings fitted a quadratic model, where experiencing a lot of either low mental workload or high mental workload significantly contributed to increased stress levels at the end of the day. Mid-range levels, which could represent either a lot of time spent at a medium mental workload level or a lot of fluctuations that could balance out to levels around a 3, were shown to be associated with less stress than the more extreme levels.

These findings were very similar for mood, where participants qualitatively expressed that too much high mental workload was associated with a more negative mood; decreased enjoyment and decreased satisfaction, two of the negative associations with too much low mental workload, could sensibly be assumed to be associated with more negative moods. Quantitatively, a quadratic model also showed that daily mental workload levels significantly contributed to perceived mood at the end of the day, where low and high mental workload levels were both associated with more negative moods compared to mid-range levels.

Fatigue was identified in the qualitative findings as another factor that is negatively impacted by a lot of high mental workload. The quantitative data supported this by finding a linearly positive relationship between daily mental workload levels and fatigue levels at the end of the day.

Additionally, quantitatively, mental workload levels were also found to linearly contribute craving as a physiological state, where higher daily mental workload levels were associated with higher cravings at the end of the day. This finding did not emerge from the qualitative analysis.

Interestingly, these four findings relating to stress, negative mood, fatigue, and an aspect of food craving, had somewhat previously been linked to mental workload levels [70; 101; 181; 190; 8]. With the exception of a fatigue study [101], which our findings supported, however, previous research has been limited in terms of quantity, and seemingly constrained to laboratory environments, measuring the effect of mental workload only over the period of work tasks, and analysing only quantitative data. This research broadened this view to the impact that mental workload levels have over a period of a day, qualitatively and quantitatively. It quantitatively collected messy real-world data, and yet relationships between variables still emerged, suggesting that we can be confident in the findings that did emerge. The quadratic curve for the stress and mood data indicates extended amounts of low mental workload has a similarly negative impact as extended amounts of high mental workload. Mental workload was not found to contribute to sleep in terms of quality, duration, or onset. Sleep duration could be sensibly assumed to be affected by life commitments, such as waking for work, and thus it is unsurprising that a relationship between mental workload and sleep duration was not identified in this study. Previous research by Goel et al [101] did find that mental workload levels during the day predicted sleep onset at night. Sleep in that research was measured objectively, and thus it could be hypothesised that there were vast mis-recollections in regards to subjectively inputting data relating to sleep onset and quality in our study. Indeed, subjective sleep recollections commonly do not match up with objective measurements [29].

Taking the qualitative and quantitative findings from this chapter together, we have strong indications of the impact that mental workload has on aspects of our wellbeing and performance. Our early data does seem comparable to physical activity, in the sense that if you don't exercise enough, or you exercise too much, there can be negative implications for your health, but striking the right balance of physical activity has endless health benefits [222].

Speculatively, once we have access to such brain data as personal data, habit formation may be interesting to study [210]. Notably in this chapter, we saw that people had goals to actively break up high mental workload periods, and designing technology to help people measure or recognise the impact of break taking on subsequent mental workload or productivity could be beneficial. Equally, as with people living a sedentary physical lifestyle, technology could help people to comprehend the scale of their prolonged low mental workload periods so they can work towards improving their mental workload activity. A common concern in research at the moment is the impact of mobile phones on e.g. mental health and sleep [173], or ability to work or study because of their distractable nature [71; 64]. In this regard, David et al [64] found that the amount of time spent social media and texting was positively correlated with students' ability to study; Douglas et al [71] found that students who spent more time distracted on their phones during lectures had lower grades compared to students who spent less time on their phones.

In terms of mental workload, it is well-known that the state of underload can lead to performance errors because of distraction [240; 204]. Our qualitative findings that looked more holistically at low, medium, and high levels, found that level of distractibility changes with level of mental workload (higher mental workload was associated with lower distractibility and vice verse). This finding was supported with the quantitative analysis, which showed mental workload could somewhat predict level of phone usage in terms of social media and total usage time. In the context of the findings from Study 1, this suggests that performance on lower mental workload tasks could be negatively affected. This is because if an individual is not able to multitask effectively when dealing with a distraction, they may either stop engaging with the primary task (which could, for example, mean they miss important information or take longer to complete the task) or make mistakes on the primary task due to not managing with the demands of both tasks. In daily life, this might not always have large implications, such as whilst watching TV, but in the same token could mean poorer performance on more important tasks such as making an online payment.

Thus, mental workload tracking pBCI technology may enable people to track the effects of activities on their behaviours, e.g. the impact that distractions have on their mental workload and subsequent ability to carry out their activities. Whilst there is much more we need to understand before detailed specific, individual mental workload goals can be determined, the results so far show the nature of the goals we could set and how they may contribute to life improvements.

5.5.2 One mental workload size does not fit all

As the primary contribution from this chapter outlined an apparent Cycle, where we ideally vary our day-to-day mental workload in particular patterns, it enabled the identification of the types of goals we should set in our 'mental workload lives' (outlined above). We generally know what is healthy for everyone in terms of physical activity; e.g. walking 10,000 steps each day is, even if an oversimplified goal, good for us. But it is clear that participants had different preferences in their mental workload lives, such as those who regarded low mental workload as positive vs those who regarded it as negative. It should therefore be considered that the 'right' amount of mental workload might differ between people. We saw that some participants perceived high mental workload as overwhelming, and so these individuals might benefit from less high mental workload fluctuations compared to those participants who perceived high mental workload as exciting. So as cognitive activity is not tangible like physical activity, there is added complexity in tracking data and future BCI technology that passively tracks cognitive activity for use as personal informatics will only suit the needs of all users if the preferences for each individual are taken into account. As some research has noted though, keeping generally active is better for cognition and cognitive health with ageing [88; 87], and we speculate that this would also be true in terms of mental workload; older participants were not the focus of the current study, but investigating how mental workload tracking could be used to avoid a cognitively sedentary lifestyle in this population would be an interesting area for future research.

One insight from our work further emphasises the 'subjective' mental workload experiences. Much mental workload research relies on the subjective reports of participants, and our work explicates further a well established principle that this is individual, and different people's experience of the same demand may vary dramatically, even for themselves depending on their recent mental workload levels. Maior et al [153] reported anecdotal evidence that some people found the same air traffic control demand stressful and difficult, while others reported it as challenging and fun. Our second theme expanded on these differences, and perhaps sense of pressure, for example, should be an element that is also captured to better interpret mental workload ratings.

Mental workload ratings are often used as the 'ground truth' for labelling states for machine learning, for classifying mental workload state according to physiological data. The variation between people, and between the experiences of the same work on different days by the same person, emphasises the challenge machine learning mental workload, and would strengthen the reasoning as to why it often achieves low classification accuracy for mental workload tasks [27]. Indeed, the consumer technology that is available tries to apply generalised initial machine learning models to work well for all users, before learning more data from the individual user. More importantly, though, the contextual experience of mental workload highlights the challenge of taking many examples of a same subjective rating even from the same person, and presuming that the same physiological response levels will be present. These factors highlight Sharple's recommendations [203] that to understand workload, we (and any consumer neurotechnology) need to understand a lot more about the 'whole system' that impacts a given moments experience of workload, rather than focusing purely on the relationship between controlled task demand and resources needed to achieve it.

5.5.3 Limitations and Future Research

Although we currently lack the exact wearable devices to measure mental workload longitudinally in everyday life, this research has contributed to understanding the nature of how tracking mental workload data could be useful as a form of personal data. However, the study was initial and exploratory; much more research is needed to build upon these findings.

The participants selected for this study were office workers as we believed this sample would be likely to have mental workload variety in their lives due to work being cognitively based rather than physical. We also presumed that those with mental workload variety in their lives would be most inclined to track that data as personal informatics. The qualitative contextualisation findings did show that participants were unanimously keen on tracking this data for personal improvements to their work and wellbeing. It would be useful to investigate whether similar themes emerged from other types of worker, and whether similar improvements could speculatively be made from tracking mental workload. Similarly, office workers as a sample was very broad, so research could look more narrowly into mental workload within different office-work professions.

For an IPA study, our sample size was considered large [207] which might have sacrificed some richness of individual accounts. We saw benefit in transferring each participant who had previously experienced mental workload tracking to the qualitative phase as they each carried over an unusual insight into their mental workload experiences, and we aimed to remain as idiographic as possible.

The Mental Workload Cycle appears to reveal a lot about the impact of mental workload in our lives, but more research is needed to develop understanding at a finer level. An arguably common consensus is that having lots of high mental workload in our lives is ‘good’ in terms of work output; our findings did somewhat validate that high mental workload does improve work output in terms of quality and quantity. But we lack understanding about how much high mental workload (and low and medium mental workload) would benefit us before they begin to negatively impact our wellbeing and work. Research into the length that individuals should sustain each level for and the amount of fluctuations that is healthy to incorporate into each day would provide a better understanding of what we should aim towards in terms of mental workload.

From the qualitative findings, it was also not established which level of mental workload participants sought out after low mental workload or if a specific one is the most beneficial; it was clear that higher levels were pursued, but not whether these levels tended to be high or medium mental workload. The quantitative findings showed that both medium and high levels were transitioned to after a low level, but as mentioned, these results did not always align with the preferences for transitions expressed qualitatively. And though medium mental workload was consistently described as sustainable, were breaks still needed at that level? Theme 4 also perhaps opened up more questions than answers, in the sense that it tells us the circumstances in which people are not able to fluctuate in their Cycles, but it does not answer what can be done to mitigate the effects of this. Finally, when considering the design of mental workload trackers, it is important to consider the different perceptions of mental workload between individuals, in terms of how some consider (e.g.) low mental workload as positive, and other consider it negatively. In these instances, the ways that data is presented to users could benefit from differing between types of user; for example, providing more positive reinforcement when people with negative perceptions incorporate the level into their Cycle.

More specifically in terms of the quantitative data, it is important to outline that the analysis was very exploratory in nature. Although statistical tests were performed and some conclusions were drawn, analysis relating to such an uncontrolled study should not be considered robust. In some sense, especially in relation to to the questionnaire analysis, relationships between mental workload and other variables emerged, and were somewhat supported by previous research; thus, perhaps here we can more confidently draw the conclusion that having ‘too much’ of both low and high mental workload in a day can contribute to negative feelings and perceptions in the evening.

The way that mental workload was quantified into one value that represented the period of time over several ratings, was by calculating a weighted average. Whilst this seems effective for gathering a general understanding of how much high or low mental workload people had in their day, the downside of this method was that it did not capture the fluctuations between levels that participants had experienced. For example, one participant may have fluctuated between levels 4 and 2 all day, and another participant might have stayed at a 3 all day; in this case, they would have had very different experiences, but their weighted averaged value would be the same. Therefore, as our qualitative findings showed the importance of fluctuations, calculating a weighted average was not able to consider this, and thus findings may have differed between participants who fluctuated and participants who did not.

The size of the R squared values for the mixed model analysis should also be acknowledged, which ranged between 0.02 and 0.11. These values do not explain a huge amount of variance in the data. However, the significant findings from the model outputs indicate a reliable relationship between the variables regardless of size, meaning that mental workload was deemed to be a contributing factor to the questionnaire and phone usage output. As mentioned, real world data is messy, and it is intuitive to recognise that many factors alongside daily mental workload levels contribute to certain states, such as stress. In regard to phone usage time, it cannot be expected that going on one's phone is the only method of distraction; it could be hypothesised that more variance in the data could have been explained by mental workload level if all form of distracting activities were incorporated into the model, such as laptop social media time or time spent daydreaming. The significant relationships, then, suggest that mental workload is a contributing factor towards the data measured in the outcome variables.

A final point regarding the quantitative analysis is that, as mentioned, the mental workload ratings app captured an overall level of mental workload over 30 minute or 1 hour intervals (as well as longer intervals when participants did not enter their rating as soon as prompted). Therefore, if fluctuations occurred within those intervals, they will have been missed. As this study was exploratory, the data collected in regards to the ratings app provided a broad insight into the types of levels and fluctuations participants experienced in their days. Future research could explore this on a more granular level by asking participants to enter their mental workload rating every time their mental workload level changes as opposed to at regular time intervals.

It will be exciting to see further research progress the understanding of mental workload from a life perspective so that we can further develop our knowledge about how the data can be used to optimise areas of our lives. We expect future work, when more generalised mental activity tracking devices are available in practice, to unravel a lot about lived mental workload experiences [196; 82], especially in relation to other devices in a quantified self ecosystem.

5.5.4 Conclusion

With the bloom of consumer pBCI technology on the horizon, it is important to establish how the data can be used to facilitate life improvements. This chapter moved away from considering just short instances of mental workload within a task and provided detailed insights into lived experiences of mental workload. Findings suggest that considering mental workload from a holistic and person-orientated perspective is important for understanding aspects of our wellbeing and task performances. Based on an apparent Mental Workload Cycle, healthy and efficient outcomes come from aiming to fluctuate between mental workload levels in particular patterns, as this prevents the negative implications resulting from sustaining any level for too long whilst enabling the positive implications that each level can provide. However, actual behaviour often appears not to follow desired behaviour, and negative implications are likely to result from this if not mitigated. Whilst more research is needed, an understanding of the nature of goals we can set in terms of mental workload has been developed. By taking into account people's perceptions and the factors which affect their mental workload ability, this study strongly suggests that tracking mental workload data is not just useful to measure during isolated work tasks.

Chapter 6

Design Perceptions

6.1 Introduction

This chapter aims to address research question (c):

How can objective mental workload tracking data be meaningfully communicated to users?

As previously outlined, the development of consumer pBCI devices is progressing rapidly. Technologically inferring cognitive activity from brain imaging measures is progressing towards a machine learning challenge, as physical activity tracking once did [236]. Just as importantly, however, is designing meaningful and safe interactions with these systems. In the previous chapter, we addressed how mental workload data could provide useful insights for people as a form of personal informatics [145; 82], where tracking, reflecting, and setting goals in terms of individual Mental Workload Cycles might facilitate improvements to certain aspects of welling and daily performances. As well as addressing 'what' type of data could be useful for meaningful interactions, there is a need to address 'how' meaningful and safe interactions between people and pBCI devices can be facilitated. Therefore, the next two chapters regard qualitative research into the perceptions of potential users of pBCI devices about what should be considered in terms of designs and ethics during the development of these devices. The findings in relation to communication of data are outlined in this chapter, followed by the findings relating to ethical perceptions in Chapter 7.

In terms of meaningfully communicating data, as discussed in more detail in Section 2.2.2, in order to meaningfully present personal informatics data to users, there is a need to understand domain-specific insights that are of particular interest to users [49]. If done effectively, visualisations can be a powerful tool for helping people gain insights into their behaviour, which encourages long-term tracking [84]. Thus, we investigated how potential pBCI users consider mental workload such that design recommendations could be produced based on these insights.

6.2 Findings

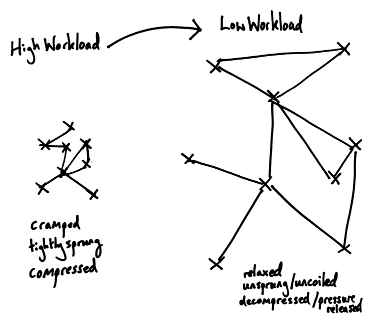
Below we present two sets of themes, first those relating to metaphors, followed by themes relating to colour and shape.

6.2.1 Metaphors

Four themes emerged regarding the metaphors our participants used to describe mental workload. These included intensity, sustainability, balance, and capacity.

Intensity

This theme captures metaphors that related to the degree of intensity between mental workload levels (low, medium, and high). For example, Participant 1 considered mental workload in terms of a sprung coil: *"For low workload I feel like mentally relaxed or like unsprung, like I had been like coiled up to do what I was doing before. Not that I hate the high workload but [it] feels like pressured or compressed or tightly sprung. I didn't use that in a negative way being like tightly coiled ... It's the idea of the pressure and then the release."* (P1, Figure 6.1a). Similarly, another example of mental workload metaphors relating to intensity is provided by Participant 15 (see also, Figure 6.1b): *"You know in the old little kids cartoons you used to watch when someone's working really hard you see the gears grinding away in their head and it gets to a super high temperature and starts steaming? That's probably like a high mental workload, and a lower one is like not much really happening - the cogs aren't really turning as quick kind of thing."*



(a) Metaphor from Participant 1 showing a tightly (high mental workload) and loosely (low mental workload) sprung coil.

(b) Metaphor from Participant 15 showing gears grinding with temperature element.

Figure 6.1: Sketches relating to Intensity as a theme.

So Participant 15 imagines mental workload in the context of a cartoon, where the intensity of the cogs turning in a character's mind relates to the mental workload level. From the two examples given above from Participants 1 and 15, mental workload is described, perhaps as expected, as relating to intensity. Interestingly, however, they are described in completely different contexts that show added complexity than simply high levels. Participant 1's sketch involves a network of activity, and that intensity is related more to tightness perhaps related to tension (e.g. tensing and relaxing of muscles), rather than something of a larger size.

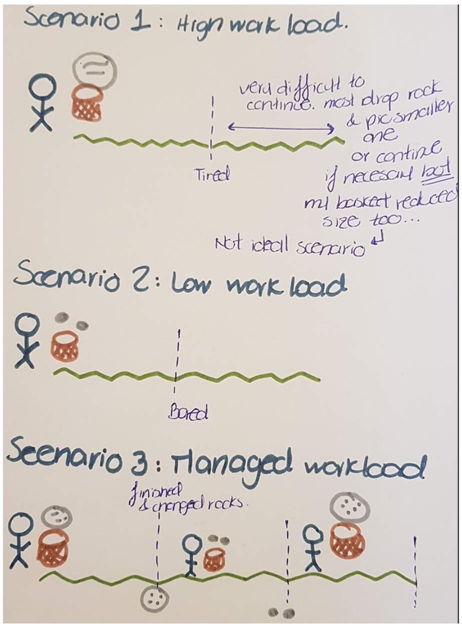
Sustainability

Many participants described mental workload in terms of metaphors relating to the ability to sustain each level. In all the sustainability metaphors, the ability to sustain the mental workload level decreased as the mental workload level increased. Interestingly, most metaphors that related to sustainability also related to intensity, and most metaphors relating to intensity also related to sustainability (except the two metaphors outlined in the first theme). In this vein, the less sustainable equalled the more intense, and vice verse, and the less intense equalled the more sustainable, and vice verse. Participant 6 provided a metaphor relating only to sustainability:

Responder: *"I relate it to as if every task had a weight, as if it was heavy or light weighted."*

Interviewer: *"Can you tell me more about the weights?"*

Res: *"So if every task was like a rock and I need to fit it in a basket, a low mental workload task would be a small rock and a high mental workload would be very big. And it's as if my mental capacity was one of these nests or a bag or something, and there were several tasks*



(a) Metaphor from Participant 6 showing basket with varying weights.

(b) Metaphor from Participant 9 showing the journey up a mountain.

Figure 6.2: Sketches relating to Sustainability as a theme.

that I could fit in a day. And I guess the size of my basket changes according to how far I've been walking with that basket or if I've slept properly."

Int: *"So does the basket represent time or your mental capacity?"*

Res: *"I think it represents my mental capacity. Yes, I think the basket represents my mental capacity. If I can extend my metaphor, I could be walking with my basket [and] there might be short distances that I could carry a heavy basket and maybe there would be longer distances that I could carry with a medium weighted basket."* (P6, Figure 6.2a)

Participant 6 above described mental workload in relation to carrying weighted objects – the heavier the object, the less distance one can walk with it, and the bigger the object, the less objects one can carry. In other words, sustainability decreases when the mental workload level increases.

The next two metaphor examples regard metaphors that fit into both the intensity and sustainability themes, as described above. Participant 9 compared mental workload to walking up a mountain (Figure 6.2b): *"[it] could be like a high lake or a high mountain. So maybe a low level would be like the lake at the base of the mountain, and high mental workload at the top, and the medium maybe the way up the mountain ... You don't need to walk or anything in the lake, you just get a bath and relax. It is harder to get to the top, so [the] beginning of the way is easier, but then you feel more tired when you are getting to the top, it's like you need to focus more to get there, so that will be the high workload for me. And the medium is like the beginning of the walk, the beginning of the way."* Participant 9 described mental workload in terms of the journey up a mountain, where the increase in mental workload correlates to increased intensity and decreased sustainability. We have somewhat interpreted the link to decreased sustainability, making this assumption due to the increasing fatigue experienced at a high mental workload level, which is unlikely to be sustainable. Interestingly, after each participant had produced their own metaphor, pre-established metaphors based on the research by Wilson et al. [237] were put to participants to establish whether any particularly resonated. One of these metaphors regarded a mountain, where the steepness of the walk reflected the mental workload level, in the same way to what Participant 9 described. 11 participants named the mountain as a metaphor that they found to resonate with their mental workload views.

A more explicitly linked metaphor to intensity and sustainability is provided by Participant 4: *"High [mental workload] would be a sprint, a medium would be like a reasonable 5k or jog, and a low would be like walking to the park or having a hike ... Which I think might be the reason why I define it in terms of sustained effort because you can't maintain a* *sprint. If you can maintain a sprint more than some amount of time then by definition it is not a sprint anymore."* Participant 4 described mental workload in terms of running, where there increase in intensity resulted in a decrease in sustainability. A number of other participants also used running metaphors to describe mental workload in this manner. A final metaphor example is shown from Participant 2: *"I'd guess it would almost be like a fuel tank for the day ... At the start of each day, you start with a full tank and similar to a car, where if you put your foot down all the time you use that fuel up quicker, so if you do more intense activities throughout the day you will use it up quicker and then you'll have less left over for later on in the day."*

Balance

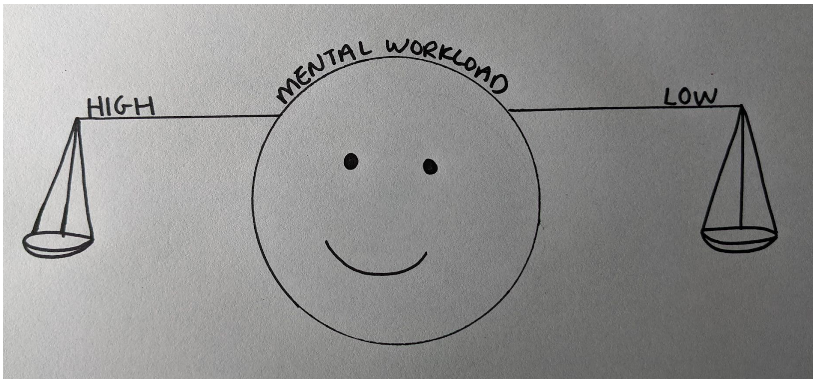


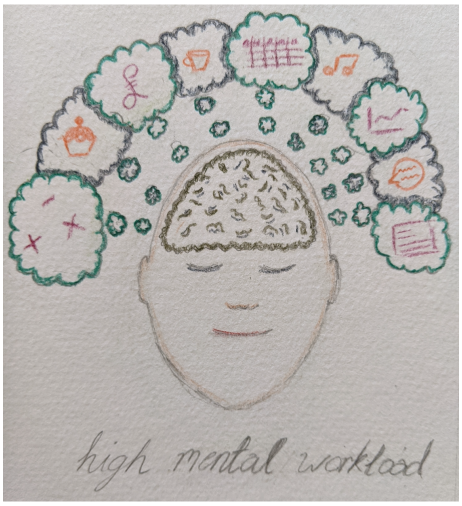
Figure 6.3: Metaphor from Participant 13 showing weights to denote balance.

Another way that participants described mental workload using metaphors related to balance. This related to finding an equilibrium between mental workload levels in order to achieve positive outcomes, and Participant 13 provided an example of this: *"I think it's like scales, I'm gonna use scales. You want the right amount of balance between high workload and low workload - too much of high [you] burn yourself out, too much of low you might just become not very motivated and freeze and not like achieve much. Yeah,* *I feel like it's scales that you want to be balanced."* (see also, Figure 6.3). Participant 13 used their metaphor to highlight the importance of balancing mental workload levels to avoid the negative effects that accompany sustaining low and high mental workload levels for too long.

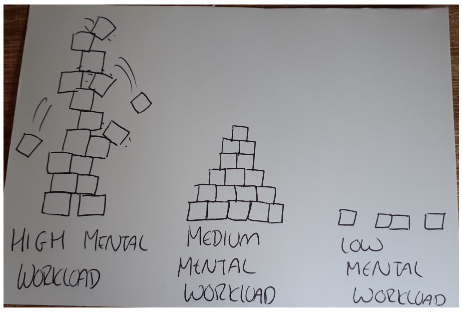
Similarly, Participant 14 used a chess example to describe the balance of mental workload levels: *"You know when people play chess and they've got that wee box that you pat, so when you've made your move and then it's the other person's turn, kinda like that but between [mental workload] levels. So you're really busy [at a high mental workload] and then you would like press a button and then you would have a break or you would be on a low level and then something happens and the button would get pressed and then it would be your [high mental workload] turn again."* In this more abstract example above, Participant 14 used the context of chess to describe the fluctuations between different mental workload levels that are experienced throughout the day. From this theme, we can see participants using metaphors to describe the balance between different mental workload levels.

Capacity

This theme captures metaphors relating to capacity that participants used to describe mental workload. In this regard, an increase in mental workload was represented by a decrease in capacity. This is demonstrated by Participant 10: *"Holding a lot of things in your head at one time and still being able to function ... It's jumbled up, there's lots of images, lots of different images, and it's work, and it's home, and it's friends, and it's family, and it's life, and it's chores, and it's everything at once but it's all of a jumble ... For low [mental workload] there are more distractions popping up all the time, but they go away quite quickly, it's less chaotic."* (Figure 6.4a). The passage from Participant 10 shows a metaphor relating to the quantity of 'things' that someone is processing. High mental workload was associated with a high quantity of chaotic thoughts or tasks, and low mental workload was associated with more available capacity, represented by less chaos and perhaps less significant items to process.



(a) Metaphor from Participant 10 showing a high mental workload state with many things represented in the mind.



(b) Metaphor from Participant 18 showing building blocks representing the mental workload levels.

Figure 6.4: Sketches relating to the Capacity theme.

Participant 18 also used a metaphor relating to capacity: *"I think it's probably like building blocks that you're trying to balance and if you have too many then you start to get nervous and worried about it, whereas if you've got a medium amount then you're like, "Yep, my tower's pretty sturdy," and if you don't have very many you're probably a bit sad cause you don't have enough blocks to build a wall - it's not very productive just having a block, that's not making a wall."* (Figure 6.4b). Participant 18 used the context of building blocks to describe how the higher the mental workload, the more blocks there are which get harder to manage. Interestingly, a medium amount of building blocks was considered an ideal quantity to create a desired outcome, but a small amount of blocks (representing low mental workload) was considered unproductive, and a large amount of blocks (for high mental workload) was considered unmanageable.

In terms of capacity metaphors, Participant 17 demonstrates another example:

Responder: *"I think it's like filling a bucket with sand and I think there's a point of mental workload where the bucket starts both leaking sand and overflowing sand, and so it's your ability to deal with the sand that's going into that bucket in order for it to not overflow or leak. I think that's how it feels because I think often when you're operating at that higher level of mental workload it's like there's lots being demanded of you and it's your ability to deal with things quickly enough that you don't reach beyond the capacity of your mental workload."*

Interviewer: *"So am I right in thinking that when you're at a low mental workload the bucket isn't very full of sand?"*

Res: *"Yeah, it's not got anything in it, there's not very much for you to deal with."*

Int: *"So as the bucket gets fuller that represents the increase in your workload?"*

Res: *"Yep, absolutely."*

6.2.2 Colours and Shapes

Three themes were found in relation to the colours and shapes that our participants associated with mental workload.

"A colour is as strong as the impression it creates"

This theme regards the colours that participants associated with the mental workload levels (low, medium, and high). It comprises three subthemes, including: the traffic light system, green for medium, and vibrancy is key.

The Traffic Light System Perhaps as could have been expected, the colours that many participants associated with the mental workload levels were red, orange/yellow, and green, representative of traffic lights. For example, Participant 2 said: *"Literally the first thing that comes to mind is the same colours as on a traffic light."*

For a large gorup of participants, mental workload was often associated with the colour red, medium with the colour orange, and low with the colour green. The account from Participant 4 demonstrates this: *"I would say I feel I'm not being very original but red [for high mental workload] . . . Orange [for low mental workload] . . . Green [for low mental workload]."* Participant 5 further illustrates this:

"It's difficult to imagine it as something other than red. But yeah, it feels so biased by the society we live in, so I can only imagine it from the normal spectrum going from green to red through some colours".

From the passages above it is clear that participants commonly associated the mental workload levels with colours representative of a traffic light. But this wasnt always the case.

Green for Medium Another subtheme that emerged from the data was that many participants described their association with a medium mental workload level with the colour green. Participant 7 outlines this simply:

"I would say medium is green." Participant 6 shared the same view, and indicates the reason behind their choice: *"I think green is a good colour for medium mental workload ... Green is peaceful."* Perhaps interestingly, Participant 3 expanded this explicity when they associated the colour green with a medium mental workload level, but incorporated that within the traffic light colours outlined in the previous subtheme: *"I would say red is high [mental workload] ... Low would be yellow ... Green [would be medium mental workload]."* (P3)

The passages above demonstrate that medium mental workload was the most desirable state to be in (not bored or overworked) and so the colour green was associated with this rather than low levels. This somewhat contrasts to the participants that defaulted to a traffic light system.

Vibrancy is Key In comparison to focusing on choice of colour, participants consistently discussed colour preferences in terms of *brightness*. High mental workload was often regarded as dark or bold, medium mental workload was bright, and low mental workload was pale or muted. Some participants outlined how they would choose one colour at different levels of vibrancy depending on the mental workload level, and other participants outlined different colours of different vibrancies for different mental workload levels. An example of selecting one colour at different vibrancies is shown by Participant 11: *"I was thinking that because I perceive mental workload as like on a spectrum, I think about mental workload [as] only one colour but like different shades of the colour. So the colour that I associate [with mental workload] is blue, and I think it will go [from] like the blue you have in the background in the sky, that would be like [a] low [mental workload] level, very close to white. And it will just increase until it will get [to a] very dark blue or like navy blue."*

From the passage above we see that Participant 11 associates the colour blue with mental workload, and the vibrancy of the blue changes depending on the mental workload level, where a pale blue is associated with a low level which gets brighter until it darkens at a high mental workload level. Participant 1 also discusses one colour in terms of vibrancy: *"I think [low mental workload is] more muted colours, like grey. I'd just pick any colour and make it like pale and muted - it's not a vibrant colour to be at a low workload in my opinion ... Whereas at the other end of the spectrum is like brighter more bold colours."*

As well as participants that associated mental workload levels with one colour at different vibrancies, Participant 16 provides an example of using different colours depending on the mental workload level: *"I think grey, muddy brown, possibly black [for high mental workload] . . . I think white [for low mental workload] . . . Medium [mental workload], I think perhaps green."* Participant 16 associated different colours with different mental workload levels, but similarly to Participant 11, the colours selected were pale for low mental workload, brighter for medium, and darker for high levels. This appeared to be a trend for participants that selected different colours depending on the mental workload level. Participant 8 provides another example of this: *"Black and red [for high mental workload] . . . Yellow and blue [for low mental workload] . . . Green [for medium mental workload]."* Although Participant 8, similarly to other participants, did not specify the vibrancy of the colours that they selected, the colours outlined did seem in line with the impression that high mental workload was associated with darker or more bold colours, low mental workload was associated with more pale or muted colours, and medium mental workload was associated with brighter vibrancies.

"I Paint with Shapes"

Two subthemes were identified that concerned the shapes that participants associated with the different mental workload levels (high, medium, and low). These included: smooth to spiky, and nature.

Smooth to Spiky Participants frequently associated 'smooth' shapes for a low mental workload level and 'spiky' shapes for a high mental workload level. Participant 5 outlines this: *"I would say that low mental workload is a smoother shape, whatever that is, you know, [a] circle, sphere, it's smoother, and the more difficult it gets it starts being less symmetrical - less smooth and more spiky."* Similarly, Participant 14 outlines this again in terms of a high mental workload level: *"It [high mental workload] would probably be a kind of triangle shape. I feel like it would be a triangle because it would be pointy."* Participant 14 again places less emphasis on the shape that they chose, instead outlining how it is the pointy characteristic of the shape that is the important factor. Participant 10 outlines this further: *"[High mental workload] would probably be quite spiky."*

For the high mental workload level, participants rarely focussed on specific shapes, but instead consistently identified the level as involving shapes with spiky characteristics. For a low mental workload, however, participants did often specify shapes, but these shapes had smoother characteristics, or as Participant 10 outlined: *"Very rounded shapes."* Participant 14 provides an example of a specific shape: *"It [low mental workload] would be a circle, just like a never-ending line."* While Participants 5, 10, and 14 opted for more circular shapes, other participants opted for flat lines. For example, Participant 9 outlined: *"I will describe it [low mental workload] as a line without too many figures or anything."*

The passage above from Participant 9 still shows the smoother characteristics associated with being at a low mental workload level with the absence of any sharp corners of angles, but is expressed in terms of a straight line instead of with rounded shapes. Participant 4 also described this: *"[Low mental workload is] a flat line."*

Nature Shapes associated with nature was another subtheme that emerged. Participant 13 incorporated the smooth and spiky characteristics as described above in terms of sea waves: *"Waves, like starting off really [participant mimics smooth waves] and then getting really like choppy."* So Participant 13 implied that being at a low mental workload level was associated with smooth waves, whilst a high mental workload level was associated with more pointy waves. Whilst this is links to the previous subtheme, many participants associated a star shape with mental workload, and this associated varied between medium and high levels, and also as mental workload as a general concept. For example, Participant 8 outlined in relation to mental workload as a general concept: *"The first shape that came to my mind was a star, I don't know why."*

Participant 8 above represents several accounts which associated a star with an aspect of mental workload, but without depth of conscious reasoning. Participant 14 outlined how a star shape was not as severe as a pointy triangle, which is why they selected a star for a medium mental workload level: *"It [medium mental workload] would be like a star probably. It wouldn't be quite as severe as a triangle."*

The passages above represent a common theme across the data where participants associated aspects of mental workload with shapes relating to nature, in particular a star shape. For the participants that associated stars with mental workload, the associations varied between medium and high levels, and mental workload as a general concept.

"Life is all about perception. Positive versus negative."

The final theme relating to the colours and shapes that our participants associated with mental workload surrounds the way in which the colours and shapes associated with mental workload can change depending on participants' experience of their activity at that mental workload level. This was described in different contexts, including colours and shapes at a high mental workload level and low mental workload level. Participant 18 describes how they associate different objects and colours with a high mental workload level depending if they are having a positive or negative experience: *"It [high mental workload] depends on the context. Some spiky red object when it's bad, probably looks a bit like a Covid, it's pretty angry because that's when everyone's at you and you're feeling prickly and just stressful. Whereas when you're doing high mental workload and you're sort of in flow and everything's brilliant, you're a genius, and you feel like you're doing amazing work, then that's more of like a swirl I suppose it's probably blue."*

Participant 18 above provides insight into how there can be different experiences at the same mental workload level, and how this can be associated with different shapes and colours. Participant 1 describes similar associations in terms of colours: *"I want to say both red and green [for high mental workload]. Green because it's a positive thing to be at a high workload, like learning and probably achieving something as well, but red in terms of there's like a speed dial getting near the top end and capacity would be red."* Participant 1 describes different colours depending on how close to the high mental workload limit they feel. They associate green as a positive colour when the high mental workload remains within their capabilities, but when the high mental workload is almost exceeding their capacity, they associate this with the colour red.

For a low mental workload, Participant 6 describes how they associate different colours and shapes with the level depending on their experience: *"I think low mental workload, depending [on] if it's a sense of being relaxed and just like resting, I do like the colour of sky blue for that and a drop of water or maybe a fountain, or something to do with water. And I think when it comes to boredom, maybe purple could work and I think I could associate it with the night."* So the passage from Participant 6 describes how being at a low mental workload level can either feel relaxed or boring, and different colours and shapes are therefore associated with the level depending on the experience. From this theme, it is apparent that the colours and shapes that our participants associated with different mental workload levels could vary, and their associations depended on whether their perceived experience was regarded as negative or positive.

Resonating Metaphors

In terms of the pre-established metaphors that were described to participants based on the research by Wilson et al. [237], the clear favourite resonating metaphor was the mountain (named by 11 participants), followed by the bubbles metaphor (6 participants). 3 participants specifically mentioned dislike of the thermometer metaphor. The passages below outline their reasons why:

"The issue with a thermometer is that the most enjoyable mental workload is when you're having fun [at a low mental workload level] and you're not working. Zero degrees is not the most enjoyable temperature, so I find it to

be a bit of a flawed metaphor." (P4)

"I feel like the thermometer one would give you numbers, but I feel it's lacking time and the spectrum, that is something that changes I think." (P6)

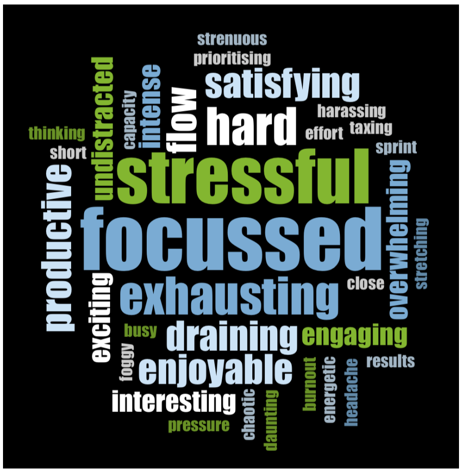
"The thermometer is perhaps a bit too linear for my liking because I think it's a bit more complex than just on scale." (P2)

The passages above highlight the reasons that the thermometer metaphor specifically did not resonate with a few participants. These included that the scale is flawed and too simplistic. The mountain metaphor resonated most with our participants, and the bubbles metaphor also resonated with several participants.

6.2.3 Word Clouds

Four word clouds were produced portraying the most common words that participants used to describe what it feels like to experience different mental workload levels and mental workload as a general concept. The bigger and bolder words represent the most frequently used descriptors.

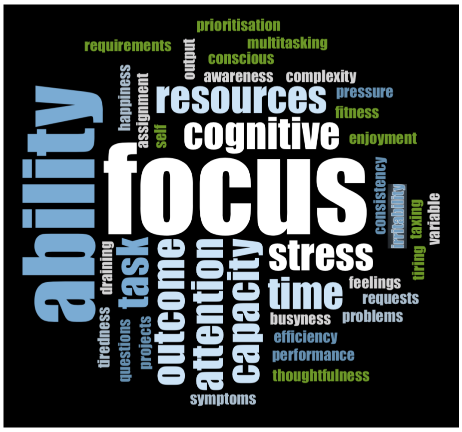
What is of note for the high (Figure 6.5a) and low (Figure 6.5c) mental workload levels is how both positive and negative descriptors are used frequently. For example, boring and relaxing were used frequently to describe how it feels to be at a low mental workload level. Boring may be interpreted as a negative descriptor, whilst relaxing may be interpreted as positive. For high mental workload, positive words such as satisfying and productive were used alongside more negative words such as draining and overwhelming. Interestingly, the frequently used descriptors for a medium



(a) [High mental workload.

(b) Medium mental workload.

(c) [Low mental workload.



(d) Overall mental workload.

Figure 6.5: Word clouds from participant descriptions of mental workload.

mental workload level (Figure 6.5b) were heavily weighted towards more positive descriptors.

6.3 Discussion and Design Recommendations

Our research was motivated by the not-unexpected and imminent arrival of consumer neurotechnology, where we will soon be able to track our cognitive activity in similar ways to which we can now ubiquitously track our physical activity. Our aim was to consider how to make such data meaningful to users in terms of data displays. To achieve this, section interviewed participants with the aim of understanding how they perceive their cognitive activity data. From building an understanding of how people conceptualise their cognitive data, and considering the maturation of personal informatics for physical activity tracking, below we present and discuss design recommendations that we hope can guide the initial design of software that accompanies consumer neurotechnology. To do this, we first characterise the key considerations and design space of personal activity tracking software in Section 6.3.1 before presenting recommendations in Section 6.3.2.

Based on research that has investigated visualisations of personal informatics data for non-experts [49; 50; 59; 120; 84; 95; 148; 81; 26; 237] to form the topics under investigation, we identified the characteristics of metaphors, colours, shapes, and descriptors that can be translated into effective interface designs that enable users to meaningfully reflect on their cognitive data.

Before outlining recommendations for design, an interesting aspect to note about the findings from this section is the consistency across the different topics. This can firstly be demonstrated from looking at the results for a medium mental workload level. For associated colours, a theme that emerged was how medium mental workload was associated with the colour green. Green tends to denote positivity and positive emotions [54] which was firmly reflected in the medium mental workload descriptors outlined in the word clouds. Medium mental workload was also found to have categorically positive associations in the personal experiences qualitative findings presented in the previous chapter.

Additionally, another theme that emerged from the colours and shapes topics was how participants visualised mental workload differently depending on their experience of the level (from *Life is all about perception. Positive versus negative*) - for example, high mental workload could be visualised differently depending on whether it was a stressful experience or a successful one. This was again represented in the word clouds for high and low mental workload, which both had positive and negative descriptors; this is also in line with the personal experiences qualitative insights that found perceptions of high and low mental workload levels could change from positive to negative (and vice verse) depending on certain factors, such as performance outcome.

We consider these to be important factors to keep in mind that should steer design away from simply implying that e.g. more activity is good or that positively associated colours should link to specific levels. Indeed as we found in the previous chapter, changing between levels, and being at optimal levels for the occasion or desired outcome, was important, and that deliberate movement through a Mental Workload Cycle is good, and that being stuck in the cycle is bad.

6.3.1 Grounding the Design Space

To ground discussion of design for personal cognitive informatics, we propose that it is helpful to first consider the maturation of design for physical activity tracking user interfaces. This is especially relevant as we consider which aspects of the themes in our findings should lead to design recommendations. Two dimensions are of note: **time** and **technology**. For **time**, we see interfaces for 1) current activity, such as a workout, 2) daily activity, which often shows both overview and the most detail across activity from the day, and 3) historic week and month long history views of activity, which aim to show trends and long term achievement. It has been noted how short-term feedback is valuable as it creates awareness of the users' current status, and long-term feedback is valuable for revealing data trends and patterns [145]. Li et al. [146] also outlined that access to personal informatics data providing details about the users' current status is important for informing users about whether they are meeting their goals or whether they need to act to correct their behaviour. The authors also outlined how a historic view is also important for establishing trends and patterns which enables users to reflect upon whether they are making progress towards their goals.

For **technology**, it is also valuable to consider the device ecology for personal informatics for physical activity trackers to ground aspects of design for cognitive activity trackers: wearables, smartphones (and tablets), and computers. On Apple devices, for example, *current activity* is captured by *wearable* technology, and personal devices (such as phones and tablets) show *daily* and *historic*[1] data for reflection. Conversely, for app usage tracking, Apple tracks behaviour on both smartphones and computers, and primarily delivers activity data through smartphone summaries in daily detail and history weekly summaries. Similarly, sleep trackers like the Oura Ring[2] capture data from dedicated hardware with no interface, and deliver daily detail and monthly history views through smartphones. It is often the *smartphone* interfaces that allow us to set goals.

It is also valuable to consider, in respect to these examples, what data is being captured and how it is being processed. Wearable physical activity tracking technology captures accelerometer/gyroscope data as well as skin sensor data (typically PPG), and processes these into recognised activities. Current workout views typically show this data as being recorded. Daily

[1] At the time of writing, Apple does not provide this data on computers.
[2] https://ouraring.com/

activity is often abstracted into a summary view, typically in relation to daily goals and historic views typically show only high level daily goal achievement per view, showing for how many days goals have been achieved. Choe et al. [50] outlined how the inclusion of immediate feedback and historic reflection is an important feature for users, but the level of detail could be improved. They argue that currently, personal informatics tools do not enable users to specify time durations on a more granular level (such as differentiating data between work and non-work hours), compare time frames, or remove outliers. For personal cognitive informatics, consumer neurotechnology will be tracking e.g. electrical activity (EEG) or blood flow (fNIRS) levels in different regions of the brain; these brain imaging methods have shown promise in their ability to track brain activity in real-world environments [187; 115]. Current consumer neurotechnology processes this data to identify equivalents to physical 'workouts', such as e.g. mediation sessions which show how often, and for how long, people have meditated; smartphone apps typically prompt people to record these sessions and help to guide their activity. It is interesting to note that, currently, most of these consumer neurotechnology devices aim to help people interrupt work, meditate, and relax, which is more in line with app usage tracking, and in opposition to physical activity tracking that encourages us to do *more* activity.

In this research, however, we are considering near-future pBCI devices that can track longitudinally, similarly to how smart phones and watches record steps and heart rate data across the day and outside of actively recorded workouts; this is opposed to just short instances of specific cognitive 'workouts'. While head-based neurotechnology might not be worn longitudinally, it is likely that more common wearable technology will be able to detect changes in cognitive concepts like stress and mental workload. Indeed, research shows that changes to stress and mental workload levels can be detected from a range of physiological responses [8] and so it is does not seem impossible that future wrist-worn technology and advances in machine learning activity recognition may infer changes in cognitive activity from the wrist, by tracking the difference between skin and blood response in comparison to an absence of movement in accelerometer and gyroscope data. Finally, where physical activity tracking converts raw data into steps, neurotechnology typically tries to classify e.g. mental workload as being either Low, Medium, or High, for periods of time and in different regions of the brain. This is a fundamental difference that moves us from counting discrete events to tracking continuous change in a set of locations over time, and means that common visualisations of physical activity data cannot necessarily translate.

Using the context outlined above, we now look at design recommendations for future mental workload tracking neurotechnology. As mentioned, the previous chapter has outlined what goals people should aim towards in their daily lives in terms of mental workload, and further research has identified that tracking mental workload in a work environment is of tremendous benefit [14; 16; 240; 204; 160]. To enable meaningful reflection on this personal data, however, visualisations need to be effective in how they display current and historic cognitive activity data to users.

6.3.2 Design Recommendations and Implications

Our metaphor themes provide four key insights about how people understand their mental workload. In particular, *sustainability* implies a key design issue that is not evident in physical activity trackers, but does have mirrors in app usage that try to help people set limits on e.g. their phone usage. We now know that sustaining any mental workload level for 'too' long has negative implications for work and wellbeing. In particular, this implies that it is not total amount of mental workload that is important to show, but what the longest sustained period is without a mental workload change. Physical activity tracking typically shows more activity as good in a day, displaying a total sum. Current mental workload activity views, therefore, should focus on how long a user has been at their current mental workload level, as an increasingly bad outcome. Similarly, daily detail views may aim to show the number of, and length of, overly sustained mental workload periods. And finally, history views could, for example, show how many days participants managed to reach their goals of e.g. avoiding sustained activity, or achieving a desired ratio of levels across the day. This principle would lend itself primarily to a view of current activity or daily detail, highlighting to user when their behaviour is undesireable.

Generally, where metaphors are an effective way of displaying personal informatics data to non-expert users [84; 95; 148], our participants were readily able to create mental workload metaphors. In particular, the mountain metaphor resonated with many participants which could be generalised further to steepness. This steepness may also lend itself to the current activity view, and combine effectively with notions of sustainability. However, metaphors of that nature most likely resonate most with users who consider mental workload in terms of *sustainability* or *intensity*. Participants also considered mental workload in terms of *balance* or *capacity*. Of these, balance may lend itself effectively to a historic view or day-detail view, showing how well a user has utilised states across the day.

Huang et al. [120] noted how users should be able to customise a visualisation enough such that they feel almost as if they have created it themselves. In these terms, it may be an effective feature to create three types of metaphor, in terms of sustainability and intensity (which can be made into one due to their overlap), balance, and capacity. The user can select their preference, depending on how they conceptualise mental workload, and thus the data can be displayed in the way that most resonates with them. Because each person is individual, the desired length that each mental workload level should be sustained for will vary, as well as the balance between different mental workload levels (Chapter 5). It seems evident that setting goals, such as an ideal ratio or balance of mental workload levels, would enable people to aim for better behaviours for them. Further, this means that if a social element to reflection is incorporated, displays might benefit from showing who has reached their goals, instead of showing the raw data as that might have different meaning for different individuals.

In terms of our colour and shape themes, there is a common societal default association that green stands for good, yellow stands for OK, and red stands for bad. The qualitative personal experience findings showed that all mental workload levels are important and good for when they are needed, to achieve balance, and that sustained activity is bad. We conclude that using traffic light colours may miss-imply that certain levels are good or bad. Instead, our findings imply that vibrancy may be a better communicator of type of mental workload. And that colour can communicate fresh mental workload periods as good, and that visualising decay over time may communicate that sustained periods are bad. Shapes may benefit from consistently remaining smooth to denote low mental workload levels, getting spikier as the level increases. These could be translated into e.g. the design of icons or even the texture of more traditional bar chart views.

6.3.3 Limitations and Future Research

Research is clear that it is important to incorporate different feedback types depending on the users' experience. While our participants were not able to experience real personal cognitive informatics, future work should study the impact of feedback on cognitive activity tracking. Findings from the personal experiences section and the word clouds in this section highlight that it should not always be assumed that a user perceives a mental workload level in a certain way, e.g. that being at a low mental workload level is always positive. This is because there can be different experiences at the same level, in the sense that someone may generally perceive low mental workload positively because it is relaxing, but they sometimes might have a boring experience at a low mental workload which is negative. In these instances, users might benefit from more positive reinforcement, such as rewards for still incorporating that level into their Mental Workload Cycle.

Future work should also further investigate ways to individualise personal cognitive informatics data for reflection. As outlined already, different people have different perceptions of mental workload in terms of generally positive or negative feelings about each level, as well as how they categorise it (from the metaphors), how long they should sustain each level for, and how they approach their daily Cycle. Personal cognitive informatics devices that track this data need to be individualised to each user in these aspects in order to meaningfully communicate data for long-term use and positive behaviour change.

6.3.4 Conclusion

Cognitive activity trackers are a realistic prospect for the near future. As well overcoming the technological challenges of tracking brain data in the real-world, further challenges lie in establishing what data is useful for users to reflect upon. In parallel with advances in consumer neurotechnology, the pressing and imperative research challenges are now how we are visualising cognitive activity data effectively for encouraging long-term use and meaningful reflections. We have identified the characteristics of metaphors that are likely to resonate with neurotechnology users, as well as colours, shapes, and descriptors. These metaphors and design associations highlight in particular the way that cognitive activity tracking will be different such that assumptions from physical activity tracking cannot be directly translated. In combination with the numerous guidelines regarding general features to make personal informatics tools effective, however, this research can inform the visual design of mental workload trackers to develop a meaningful and useful form of personal cognitive informatics.

Chapter 7

Neuroethics

7.1 Introduction

This chapter aims to address research question (d):

What should be ethically considered when developing mental workload pBCI devices, or neurotechnology in general?

Whilst there are many current and required active and ongoing discussions about ethics and regulations in the development of pBCI devices, there is a gap in research relating to the ethical concerns and perceptions held by the end users of this technology. Thus, we wished to further research into the ethical considerations of consumer pBCI devices by investigating the ethical concerns and perceptions of potential consumers. In doing so, further ethical, legal, or social considerations of neurotechnology might be established, and the already established guidelines mentioned previously in Section 2.4 might be further validated. Addressing ethical considerations at a relatively nascent stage of development is an aim in neuroethics [100].

7.2 Findings

These findings were published in the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT):

Serena Midha, Max L. Wilson, and Sarah Sharples. 2022. Ethical Concerns and Perceptions of Consumer Neurotechnology from Lived Experiences of Mental Workload Tracking. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22).

Three superordinate themes were identified (Table 7.1): 1) fear of the data, 2) the negative effect of the data on the self, and 3) the spectrum of sharing. The first thing to note about the data is that participants frequently made comparisons between this 'Fitbit for the brain' data and data collected from physical activity trackers, where the concerns expressed and points made were considered as similar or comparable. For example: *"I guess it's [tracking objective mental workload data] similar to the sense that like I wear my watch, my sports watch, literally 24/7 for the last three years that I've had it . . . you can have it track all your data and see how far you've been, see where you've gone to, the places you've travelled to and things like that. I enjoy that level of data."* (P2).

The passage above highlights that Participant 2 considers the level and type of data collected from their sports watch to be of a similar nature to the mental workload data that could be collected in daily life. This notion was apparent throughout multiple transcripts, and suggests that tracking brain data is considered a similar concept to tracking physical data.

Table 7.1: Table showing the final superordinate and subthemes from the qualitative analysis.

Superordinate Theme	Subthemes
1) Fear of the Data *Describes concerns relating to data judgement.*	Fear of personal judgement Fear of consequences Fear of inaccurate judgement
2) Negative Effect of the Data on the Self *Describes concerns relating to the negative personal effects of data tracking.*	Being controlled by the data Data exacerbating negative states
3) The Spectrum of Sharing *Describes concerns and views about data sharing.*	What concerns? Controlled sharing for positive change It depends on the risk An absolute no

7.2.1 Fear of the Data

The first theme presented regards participant's concerns about the judgements arising from the data and include the subthemes 1) fear of personal judgement, 2) fear of consequences, and 3) fear of inaccurate judgement.

Fear of Personal Judgement

Participants often reported that they were concerned about people in their lives viewing their mental workload data and making assumptions about them as individuals. This is demonstrated by Participant 2: *"I wouldn't want someone having that [personal MWL] information on like a daily or weekly basis and then them making criticisms off that basis."* (P2)

Participant 2 described their judgement concern in terms of being criticised based on their data. Participant 1 reflected on their opinion similarly in terms of their data being viewed by employers: *"It shouldn't be any concern of an employer or supervisor as to how hard I'm working if I can produce* *the results. My concern would be someone seeing it and then judging the workload based on their perception of their own workload."* (P1)

As well as describing their concern in a workplace environment, Participant 1 also outlined their concern surrounding their fear of judgement by friends and family members too: *"I don't think my friends or family need to see it cause I don't know what they'd think of the workload. They're like 'Oh a lot of high workload on Friday evenings 7-9, what are you doing?'."* (P1)

The passages above represent a common feeling of concern across the data set of being judged on a personal level by the mental workload levels that have been tracked in their lives. This applied to both a social and work environment and was perceived negatively.

Fear of Consequences

Another frequent concern from participants regarding the tracking of mental workload in their lives was the negative repercussions that might arise if their data was accessed and judged by people with authority in the workplace: *"They [the boss] could use it in the wrong way and use it as like a punishment. Like, 'You're not being very productive,' or like,'You can't cope with your new promotion,' or whatever."* (P13)

From the passage above it is apparent that Participant 13 is concerned about their data being used against them in the workplace to affect their position in the company. Participant 5 provided another example of this: *"I think I would be worried that workplaces might judge people by this sort of thing [mental workload data] and might discriminate based on that."* (P5)

As well as concerns relating to consequences in participants' places of work, Participant 15 described a situation in which the data could have negative repercussions even in the recruitment process: *"I'd have issues if it became a widely available thing because the data collected would be so accessible. Like in the scenario where I'm applying for a job and the employer asks me for my mental Fitbit data because you know, in 20 years it's just become the norm - you attach your mental Fitbit results for the last week or whatever on your CV ... I think it's not necessarily representative of how good of an employee you would be. I think a lot of people that would make great employees would miss out on a job just because their numbers aren't as high as the others."* (P15)

So Participant 15 describes how the data being accessed by employers might have negative consequences for employment opportunities, as they believe the data might be used to make negative assumptions which then have negative consequences.

As well as concerns about consequences related to the workplace, participants expressed concerns about the data collected being exploited. This can be shown in the following passage: *"They introduced those watches to track your health and all that, and it's a great idea but it took, what, maybe a month before an insurance agency used that as a way of increasing premiums on you. So there's no limit to how much those kind of tools could be exploited for other things, like someone denying you a raise cause apparently your workload is not very high."* (P4)

So here Participant 4 highlighted a data exploitation concern relating to how mental workload data could be misused against individuals in terms of discrimination; participants also expressed a data exploitation concern in terms of companies taking advantage of the data through targeted advertising: *"I'm kind of relating that to the online activity you have. Like it has happened to me that I have talked to someone like,'Oh I'm thinking about* *buying a flight,' and then it just suddenly appears the ads on Facebook and on Instagram and all the things about flights ... I don't know really why I'm worried, but I know that's the way it shouldn't be, that they have access to all the data and they can use that for their advantage."* (P8)

So whilst participants may not yet be clear how the objective brain data might be exploited exactly, it is clear from this subtheme that participants are concerned that it will be used in ways that is considered intrusive and discriminatory.

Fear of Inaccurate Judgement

Participants often reported how important context is for interpretating the mental workload data accurately, and were concerned that inaccurate assumptions might be made if the data is viewed by an external person without understanding the context. A passage from Participant 2 highlights this: *"I guess the worry would be that you get to the point of it [tracking mental workload] becoming mandatory for work and someone's regularly looking at it and analysing all your data, and then uses that as justification at work. I don't like that. I think then you'd get to the point of micromanagement and stuff ... Whilst I like having that information for myself, I wouldn't want other people to look at it and make assessments off the basis of it ... There's more to it than just a number, you know, like a number ranking or a rating, something like that. It's only half the story I guess. So it'd be useful for just you personally to have a look at, but if someone whose got no context looks at it, you know, something bad might have happened ... There's no context for it."* (P2)

Participant 2 demonstrated concerns which regarded the importance of the context of the data. Participant 3 demonstrated the same concern:

"[I would be concerned] that it [the mental workload data] would be misunderstood or misconstrued because out of context you could make some assumptions about the data that may or may not be correct." (P3)

As well as the context being important for understanding mental workload data, participants were also concerned that the data itself is complex to understand and hence can be easily misinterpreted by employers viewing the data: *"I think there is a lot of nuances with brain activity so there has to be a lot of understanding, conceptualisation and training to understand it. So I feel if we give this data to employers, to industries, they don't have the skills to understand this and I think they will make a simple use of the data; they will look for high levels,'Ok we are looking for high levels, high levels are good,' which is not true and yes that will be used against employees. I think it's a lack of proper understanding of the data, it's a lack of 100% relationship between brain data and the outcome we have in the work, and the potential of negative effects on employees."* (P11)

Here Participant 11 described a scenario in which employers have access to their employees' mental workload data yet are not equipped to interpret it correctly, which could result negative outcomes for certain employees. Similarly, concerns were also expressed regarding the effect of inaccurate data or inaccurate assumptions on those tracking the data for personal use. Participant 18 demonstrates concerns relating to inaccurate data: *"I think it [a mental workload tracker] would just need to be quite robust in terms of its science. So for example, Fitbits and lots of those devices have got cautions in them, so say the ones that take your pulse . . . they have to have threshholds so high that people who might be a bit tired cause they overdid it on a run don't just pitch up at A&E like 'I'm dying,' it's like, "No, just calm down have a glass of water." So I think this would have to be grounded in some really good sort of cognitive science to know some of the differences* *between high flow states and high anxiety states."* (p18)

Participant 3 outlines concerns relating to misinterpretations of data: *"If something as well understood as heart rate can be misconstrued in a medical setting for my benefit then I think this sort of data [mental workload] could be misconstrued."* (P3)

From the passages above we can see common concerns surrounding how devices that track mental workload data might lack validity or remain open to interpretation. Participant 13 captured these concerns in one passage: *"Maybe it's hard for an app technology to fully understand how you are. I guess it would make assumptions and I dunno, it's technology isn't it, it's an app, it's like not like, you know what I mean, you might take things too seriously. Like a Fitbit for your body, you take it too seriously like,'Oh it's telling me that my heart is permanently, I dunno, too fast,' you take that too seriously and it might make you make big life decisions based on yeah, assumptions."* (P13)

So this subtheme describes the concerns expressed by participants about judging the data inaccurately which could result in negative outcomes. From the importance of context, to the inability to make correct interpretations, and lastly, as Participant 13 outlined, basing personal decisions on data that is either inaccurate or misinterpreted, these factors are all speculated to potentially result in negative outcomes for individuals.

7.2.2 Negative Effect of the Data on the Self

This theme describes participant's concerns about the personal effects that tracking this data might have. It includes the subthemes 1) being controlled by the data and 2) data exacerbating negative states.

Being Controlled by the Data

A number of participants reported feeling concerned that the data could result in individuals becoming obsessive. Participant 10 demonstrates this: *"One [concern] is if you become too obsessed with it and it becomes a fixation and you can't stop looking and tracking. I know people have done that with heart rate monitors, they're like,'What's my heart rate now? Oh my gosh it's 72, 62,' so you know you can get obsessed with it. So finding a way to make sure that doesn't happen would be a concern."* (P10)

The passage above outlines how Participant 10 feels apprehensive about the personal impact that tracking objective mental workload data might have on individuals in terms of displaying obsessive behaviour. They refer to knowledge they have about the experiences of people they know and their relationships with their physical activity trackers and draws similar concerns for the mental workload data being discussed.

Participant 19 also discussed their concern of being controlled by the data, and similarly draws upon comparisons of physical activity trackers: *"I think you've gotta be careful because if you are relying on it too much as a validation strategy for what you're thinking then it could have an adverse impact. For example . . . it might be that 'We've noticed your mental workload's been high for a long time' . . . There is an inherent danger of relying on it . . . You wouldn't hold any reliance on a fitness app, you'd only use it for support, and the same goes for a mental workload app."* (p19)

Data Exacerbating Negative States

Participant 18 provided a rich account, again based on their knowledge of their friends' relationships with their physical activity trackers, detailing their concern that tracking this data in our everyday lives might be unhealthy for some people with mental health difficulties:

"I suppose your only concern is if you've got somebody who if they have anxiety or if they have depression or something like that, are you giving them a rod to beat themselves with? So you know how some people they have a really negative relationship with their Fitbit, they're like,'I didn't close my rings today, I'm such a fat this and I'm disgusting that and I'm never gonna this and blah blah blah blah.' I have a few friends who just use it to beat themselves and it's very hard to watch, it's very hard to stop. So I think it's about if you have high mental workload ... I think it's about how you share those messages whether or not it can be interpreted positively in a kind, reassuring way rather than,'My Fitbit says I'm having a mental breakdown,' so I think that's where your risks lie that when it flags, how do people feel when it flags? What do they do? Does it have coping mechanisms? Does it give you advice? Because otherwise you can just reinforce people and escalate their worrying about workload so then they feel like they've got more workload and less able to deal with it. I think that would be the difficult side of it to navigate." (P18)

From Participant 18's passage, we see a speculated comparison between their experiences of how physical activity trackers can exacerbate negative cognitive states, and how tracking mental workload data might result in the same difficulties amongst people with mental health problems. They outline their belief of how important it is to present the data in a way which can only be interpreted positively, instead of providing some people with *"a rod to beat themselves with."*

7.2.3 The Spectrum of Sharing

There were some stipulations as well as black and white views that arose from discussing the collection and sharing of data. The subthemes that emerged included 1) what concerns?, 2) controlled sharing for positive change, 3) it depends on the risk, and 4) an absolute no.

What Concerns?

Some participants simply had no privacy concerns about their brain data being tracked in their everyday lives: *"I'm fully aware that I've got a digital footprint that is far flung, I see no real issue with it. In fact, I'm forever selling my personal data. I'm someone that will happily do, you know, surveys for things and no doubt give too much of my personal data and information but no, no issues."* (P16)

The passage from Participant 16 reflects a number of other participants who seem to have no concerns about their data being tracked in their daily lives. Participant 7 provides another example of this: *"I don't think I have that many issues. I don't have issues - like if you could get like the Neuralink implant tomorrow, if you could volunteer for a free trial, I would be like, 'Elon [Musk] put me one, I just wanna be part of the trend.'"*(P7)

Controlled Sharing for Positive Change

In regards to sharing their data with their workplaces, participants often reported that they would do so if they were in control of who could access it: *"I think I would be prepared to share it with some people but I'd like control over who I share it with, and that would not necessarily be my boss."* (P10)

Participant 5 also demonstrated willingness to share if in control of their data: *"I wouldn't want it shared with anyone unless I give my consent ... Maybe if it allows you to have a better conversation with your manager or something like that to improve your experience or your quality of life, I think then it would be good. But again I think that should be an individual's decision."* (P5)

With several participants requiring control of who their data is shared with, the reasons for sharing their data emerged to be for the purpose of personal improvement or company improvement. Participant 8 provided an example of personal improvement: *"I know that he [supervisor] will handle the data like correctly and maybe that would help me to improve my productiveness."* (P8)

So Participant 8 would be willing to share their data to improve their personal productivity levels. Participant 16 also described how they would share their data for personal improvement in terms of reaching their potential: *"Particularly if you're being under utilised, for example, or your mental capacity is being under utilised unintentionally, then it [sharing data] might bring some benefit."* (P16)

The passages above show how participants would share their mental workload data for personal improvements. Participant 7 described the sharing of data for company improvement: *"Maybe it [sharing data with their boss] can drive the company. Especially now after the lockdown and quarantine periods, maybe they can get to know like, 'Ok this group of people are actually quite effective working fewer hours, they still get everything done,' maybe they can change the working hours or the working environments to actually benefit people in that sense. If they're really conscious and really people orientated, if they want people who aren't enjoying it or are unhappy* *and are struggling to meet deadlines, they can tell, 'Ok how can we help them? Because if we help them, we help the company,' so I like to think they would make good use of it."* (P7)

In the passage above, Participant 7 reflected on how sharing their data with the company might lead to changes of how the company operates and this might have a positive effect on the employees' lives.

It Depends on the Risk

When discussing the tracking of objective mental workload data in pilots, participants were widely more accepting of mandatory tracking in safety-critical jobs: *"I think if you're a pilot then maybe yes [mandatory tracking is acceptable] just because you have lives in your hands. It's not like you didn't submit an Excel spreadsheet that you were asked for."* (P8)

We can see that Participant 8 deemed the difference between safety-critical workers and office workers as significant for the right of employers to access their data. Participant 4 compared mental workload data to other performance checks that pilots are routinely subject to: *"It wouldn't be too dissimilar as someone checking that the pilot is not drunk before flying a plane and that would be very intrusive in many jobs, but for a pilot it's fine."* (P4)

So whilst the consensus was that mandatory objective tracking in safety-critical workers was more acceptable than office workers, it was also frequently reported that even in these areas of work, the data must not result in personal negative consequences for employees.

"I think the whole structure of how society operates around this type of high-risk jobs should change. As in not say, 'Well you're off the job and obviously

you can take less than the other pilot so then I'm either going to fire you or just going to pay for, you know, you can only work three hours today rather than six ...' So, I think obviously that would again lead to some sort of discrimination. But if the data is used to better understand these limitations at the impersonal level, so it might end up that you can see that for 95% of the pilots five hours is too much usually, so then without pointing fingers at individuals you might overall change the policy of the company to, 'Ok no one ever has to work more than five hours,' or whatever that time is so that statistically you reduce that risk ... So maybe the company should make decisions based on large data sets statistical decisions." (P5)

The passage above from Participant 5 reflects on how tracking mental workload in safety-critical jobs in the context of pilots should not be used on a personal level, but instead be used to make company improvements to improve the safety in these jobs. Therefore, whilst this subtheme outlines how the risk of the job affects how participants viewed the enforcement of compulsory tracking, it was still deemed important that the data should not be used on a personal level.

An Absolute No

Some participants were firmly unwilling to share their data with their workplaces. Participant 4 described their reasoning: *"Scientists yes [I would share my data] obviously, boss absolutely not because it shouldn't be to be shared ... It could be used for good, like you could have a good boss saying, 'Oh my gosh, my employee's always at a four [high mental workload level], I need to do something before they crack down and kill themselves,' and that would be good. However, for each good boss doing this, you would have a bad boss saying, 'You're still at a two [low mental workload level] are you* *just not working enough? I'm gonna give you more work.' As long as the task gets done the mental workload shouldn't be tracked or it shouldn't be a concern."* (P4)

From Participant 4's passage, we can see that whilst potential positive outcomes of sharing the data were acknowledged, they believe the data simply should not be shared because of the negative outcomes that have the potential to arise. This black and white unwillingness was shown by a few other participants; for example, Participant 1: *"I wouldn't want an employer to have it. I don't think they should be able to see that kind of thing for various performance reasons, like it being used for review."* (P1)

7.3 Discussion

Neurotechnology devices are starting to become available to users and literature is increasingly producing guidelines aiming to mitigate the negative implications that consumer neurotechnology will unintentionally bring [139; 73; 228; 125; 124; 216]. That research has generally operated by identifying gaps in existing ethical and legal frameworks that do not accommodate for the addition of consumer neurotechnology into the market, and discussions and guidelines are outlined in relation to this. We ran empirical research, however, which researched the views of potential end users of consumer neurotechnology; participants were not made aware of the current status of discussions and regulations, and this contributed to an uncontaminated insight into the ethical concerns and perceptions held by those who may be future end users. The aim of the research was to ground current guidelines in further evidence and investigate whether there are any further factors that should be considered in relation to the development of neurotechnology.

We hypothesised that our findings would relate to concerns regarding privacy, data validity, and personal identity, as these have been recurring concerns outlined by various authors [139; 125; 124; 216]. Indeed, these concerns were prominent in the analysis. Firstly, concerns about privacy were widespread across our findings, and issues relating to privacy are also perhaps the area given the most concern in previous research [125; 124; 216]. Specifically, as well as the explicit wish to keep their data private (from theme 3), theme 1 (*concerns relating to personal judgement, personal consequences, inaccurate data judgement*) and *sharing data if in control* (a subtheme from theme 3) all related to the concern of data privacy. This finding further validates previous research which outline major concerns relating to privacy [139; 125; 124; 216]. If privacy is regulated properly, many of the ethical concerns identified in our findings could be mitigated, enabling consumers to enjoy the many potential benefits of tracking their brain activity. Additionally, this finding is interesting as it provides insight into the daily applications of privacy concerns that consumers might have. This provides a different angle to what is commonly seen in the literature, where discussions tend to centre around how privacy should be approached (such as suggesting it should be treated in the same way as other sensitive personal data [216]), with less explicit links to experiences in daily living and the explicated implication in peoples lives.

Secondly, concerns relating to data validity have also been outlined numerously [139; 125; 124; 216]. It has been noted that a number of current consumer neurotechnologies have limited precision [124]. Our findings also outlined concerns from participants regarding the data being inaccurate, and decisions being made based on inaccurate data. This highlights the requirement for transparent and regulated claims about validity, so that consumers do not experience harm from misleading data.

Theme 2 (*negative effect of the data on the self*) related to issues surrounding personal identity. This has previously been outlined; for example, Kreitmare [139] described two guidelines relating to the self (firstly changing people's views of the world and secondly altering people's self-identity) and the IBC report [216] described how algorithms can dilute the sense of self due to helping to make a person's decisions. Our findings align with this as they regard to how the data may alter the self. The subthemes *being controlled by the data* and *data exacerbating negative states* both highlighted specific applications from participants' lives about how neurotechnology might negatively change the state of individuals. This is important if it will affect Mental Health (rather than help people as is often advertised) and again helps to ground previous research in further evidence, and suggests there needs to be regulations surrounding the presentation of data (discussed more below).

There were two findings from our data that appear novel. The first one relates to the use of neurotechnology in safety-critical jobs, where privacy was deemed less of a concern if mandatory brain tracking could increase safety. This indicates a distinction between workplaces regarding what may be acceptable for the way that data privacy is handled. Secondly, participants were concerned (*fear of inaccurate judgement*, from theme 1) that even if their data was transparent and valid, themselves or their workplaces may not interpret it correctly, especially as the context of the data is essential for its understanding. Proper regulations surrounding data privacy might again mitigate the effects of external individuals (such as workplaces) misinterpreting the data, but that does not counter the concern that the data may remain open to misinterpretation by whoever views it.

7.3.1 Current Status of Concerns

The technological progress we are seeing has been coined the 'neuro-revolution' by Ienca and Andorno [124], which is expected to follow in the footsteps of the 'genetic revolution' that reshaped some of our ethical and legal notions. Currently, however, no mandatory governance framework specifically for brain data has been established in supranational or international law [125]. The European Union's General Data Protection Regulation (GDPR) is legally binding and concerns regulations for how personal data must be handled, from collection to processing. However, even if brain data is considered as sensitive personal data, Ienca [125] noted how the GDPR in its current state leaves gaps for brain data vulnerable to breaches of privacy; they suggest regulations that consider brain data in its own category, which could protect against vulnerabilities that are unique to this type of data, as is the case already with genetic data [125; 124]. The lack of regulations around privacy for brain data are so severe that there are currently no safeguards to protect brain data from the same data-mining and privacy intruding measures that we see with other types of data [124].

In light of the finding from this research, in which participants viewed safety-critical workers as having less rights to brain privacy compared to other jobs and individuals in order to improve safety at work, this sparks a discussion about different regulation requirements for different consumers. It appears that it may be in our better interests to shape privacy regulations around circumstances, such as for those purchasing neurotechnology for personal use compared to safety-critical workplaces purchasing neurotechnology to monitor their employees. However, guidelines from the IBC Report [216] strongly recommended legislation which requires all employees to have the right to refuse the use of neurotechnology without being excluded or devalued. Our exploratory finding does suggest, however, that there is the potential that for certain situations (safety-critical), neurotechnology use could be categorised similarly to other required but intrusive measurements that safety-critical workers are subject to, such as drug tests. However, tracking brain data is complex, and if workers did provide consent for their data to be tracked, neurotechnology may access brain data that is outside of the users' awareness, meaning traditional informed consent processes may not be suitable for the use of brain data [216].

Regarding concerns about data validity identified in the current research and previous research [139; 125; 124; 216], there are stringent regulations around the use of medical devices (the EU's Medical Device Regulation, or approval from the US Food and Drug Administration), but most consumer neurotechnology companies avoid classifying their devices as medical by marketing them for wellness, relaxation, and other non-medical purposes [228]. This means that they are not subject to the stricter regulations [125] and users are not guaranteed that the data is valid and representative of true cognitive function [73; 125]. Progress in enforcing responsible innovation is being made [98] as ideas surrounding its governance are being discussed (such as the suggestion of neurotechnology developers subscribing to taking a responsible innovation oath [216]), but regulations surrounding data validation have not yet been fully established.

Concerns relating to personal identity are perhaps more of a grey area when it comes to regulation as it relates more to the characteristics of individuals as opposed to the rights that each person should have. Participants in this research drew parallels between physical activity trackers and cognitive activity trackers to describe their concerns surrounding personal identity by describing their experiences with the trackers and extending these to neurotechnology. Similar personal identity concerns have been shown to apply to physical activity trackers [140; 83], such as users feeling decreased enjoyment associated with their physical activity [83]. Neurotechnology could therefore explore the approaches taken in regard to physical activity trackers which aim to mitigate the effects of using the technology on factors relating to personal identity. It has been suggested that to mitigate compulsive, addictive, and distracted behaviour in regard to physical activity trackers, the technology could incorporate periods each day where access to quantified data becomes unavailable to users [140]. With physical activity trackers already arguably ubiquitous, however, it would be sensible for consumer neurotechnology to be attentive to the issues and solutions surrounding personal identity in the physical activity wearable field in order to account for these effects at an earlier stage of growth. Understanding these negative affects of tracking is especially important for technology that is advertised as helping mental health and wellbeing; neurotechnology could have the opposite effect especially if it is measuring cognitive activity that is directly involved in mental health conditions.

The concern relating to the misinterpretation of neurotechnology data identified in the current research appears novel and thus it is not clear whether there are active discussions in terms of identifying guidelines aimed at lessening the negative effects of this. Again it seems sensible to draw on the similar issues between consumer physical activity wearables and consumer neurotechnology, especially when considering newly established concerns. Choudhury et al. [51] outlined how misinterpreting physical activity data can negatively affect wellbeing by causing a sense of panic; this can also lead to seeking unnecessary healthcare that can put strain on health services. Due to the risk of negative implications arising from data misinterpretation, designing user-friendly interfaces has been strongly emphasised along with clear user manuals [51]. This is a simple and yet important consideration for the development of consumer neurotechnology. Indeed, an aim of our research involves investigating ways to effectively communicate mental workload brain data to users; the findings from this section will help to emphasise stringent checks for potential data misinterpretation.

7.3.2 Limitations and Future Research

This research was valuable in the sense that it enabled real-world insights on a granular level from potential users into the ethical concerns and perceptions of neurotechnology. By using this approach, we were able to provide tangible evidence supporting several concerns that have been discussed based on robust theories, and provide a different perspective to raise further ethical concerns which may not yet have been considered. Indeed, the IBC Report [216] highlights the necessity to anticipate the effects of implementing neurotechnology by using scenarios where society and future technologies are imagined and how they will interact. However, by running an empirical study, we have approached the research from an HCI perspective, which may lack understanding and detail into the depth of the topics under discussion in the ethics field.

Based on our novel finding regarding the privacy rights of safety-critical workers, this perhaps raises more questions than answers. The result was only based on a small sample of office workers, which may differ to the opinion of other samples and those safety-critical workers who would be tracked. And if safety-critical workers did consent to tracking, the effect that tracking may have on performance should be considered. Therefore, whilst the finding certainly raises an interesting point for discussion, much more research is needed before being able to establish potential legislation that has both human rights and safety maximisation at its core. It should also be noted that the type of neurotechnology under discussion in the interviews was narrow, as only mental workload trackers were considered. For the design of the study that involved interviewing participants about their experiences, it was necessary to focus on a type of neurotechnology. This narrow focus differs to what is commonly seen in literature concerning the ethics of consumer neurotechnology, which discusses issues associated with all types of consumer neurotechnology. Whilst this paper regards a certain type of consumer neurotechnology, the results should be generalisable to other consumer neurotechnology; this is supported by our overlapping findings to other research, as outlined above.

7.3.3 Conclusion

This chapter presented a novel empirical approach to understanding ethical concerns and perceptions surrounding the growth of consumer neurotechnology. To ground the interview discussions, people that had experienced tracking their own mental workload were probed about their views, which enabled insights into the concerns of potential neurotechnology end-users and examples of daily scenarios to which these concerns applied. The results relating to privacy, data validity, and personal identity provided further validation for concerns that are currently under discussion. The results relating to privacy in safety-critical jobs and misinterpretations of data highlight further important factors that should be explored further.With the introduction of mass consumer neurotechnology on the horizon,it is imperative that progress is swift to regulate its use in order to mitigate any unintended consequences and enable users to flourish.

Chapter 8

General Discussion

Consumer pBCI devices are arriving, and we are largely unprepared. Alarmingly, consumers are not yet protected through brain data regulations. People are worried that the data they are tracking is not valid [139], and their concerns cannot currently be eased by researchers in the field. Further, currently available devices focus on tracking cognitive activity over short instances of time. Arguably, however, physical activity trackers have moved on from this stage, and are now ubiquitous in day-to-day life instead. In this regard, a more longitudinal approach to physical activity monitoring is taken. Within this, people may adjust the way that they track, such that they treat the data collected from isolated workouts differently to the data collected from their wider lives (e.g. closely scrutinising heart rate data related to their workouts, but not reflecting on heart rate data outside of these workout windows). In addition, the goals set by people in regard to their data may have different facets, such as aiming to walk a certain amount of steps over the course of the day, but also aiming to burn a specific amount of calories during an isolated workout window. Crucially, physical activity trackers facilitate these different ways of tracking and treating the data by providing numerous and personalisable options, such as switching between more passive tracking and specific workout monitoring.

It is therefore plausibly proposed that these days, physical activity is generally no longer tracked just for short instances of time. Instead, people who track their workouts also track their daily data in order to build a broader, life-based understanding of their physical state. Indeed, even when we are working, talking, or eating, we are constantly producing quantifiable data for collection that could increase our understanding of our physical wellbeing and produce actionable insights for how to improve on it. And based on its widespread adoption, this data is clearly meaningful to people; technology that does not produce such meaningful insights will not be adopted by users [145; 80]. In terms of cognitive activity tracking, there is a lot of room for growth for identifying what data could provide meaningful insights to users and allow them to set goals. This can especially be considered from a longitudinal perspective, where users could track, reflect, and manage their cognitive activity in daily life in a similar way to ubiquitous physical activity tracking. Technology is advancing rapidly [14; 240] and it is speculated that we will be able to monitor cognitive activity in real-world environments as intricately as we can currently monitor physical activity. It is therefore an opportune time to consider how cognitive activity data might fit into our lives from an HCI perspective.

This narrative is the crux of this thesis, which has investigated the longitudinal tracking of cognitive activity in daily life. In this regard, mental workload was the cognitive activity of interest based on its prevalence in our daily lives and ever growing relevance. We primarily investigated mental workload in daily life as a form of personal informatics, which was a multifaceted topic.

8.1 Thesis Summary

A first step towards understanding life-tracking of mental workload data was to take measurements out of tightly controlled laboratory or safety-critical environments and into more uncontrolled environments with general life tasks. The aim of this was basically in respect to 'Can we do it?' and 'What do we need to consider?' To address this, an empirical study using fNIRS for naturalistic reading and writing tasks was conducted and was outlined in Chapter 3. From this research, challenges were identified for real-world measurements of mental workload data. These related to the sensitivity of the measures and the interpretation of the data, both of which can be situationally dependent.

While future work for Chapter 3 generally concerns challenges for machine learning research, we then considered the next big HCI question - what will tracking this data mean for our lives? How can we meaningfully gain insights? What will our goals be? This is a concern relating to personal informatics [82; 145; 194; 196]. Thus, from a qualitative and quantitative perspective, a novel longitudinal approach to mental workload was taken. Qualitative insights from the personal experiences section identified individual perceptions of mental workload, how people optimally approach mental workload in their lives and why, and factors that can be a barrier to executing this. The quantitative analysis built on this by furthering our understanding of the mental workload implications on life and the types of mental workload levels actually experienced day-to-day (Chapter 5). Chapter 6 and Chapter 7 regarded the practical development of future mental workload tracking pBCI devices, and provided guidance surrounding how the data could be meaningfully communicated to users and what should be ethically considered during their development.

8.2 Contributions and Implications

Taking the findings presented in this thesis together, this thesis could somewhat be considered as a package that improves understanding of the human (computer interaction) side of developing future mental workload personal informatics tools. Indeed, sophisticated technology systems are worthless if users cannot meaningfully interact with them in order to accomplish their desired goals [133]. In this respect, the package relates to information regarding how and why mental workload data could be useful to track, how the data could be meaningfully communicated to users, what should ethically be considered in the development of tracking tools, and challenges for real-world measurements of mental workload data.

8.2.1 Contribution 1 - Measurement considerations for general life tasks

More specifically, the first contribution identified challenges for measuring mental workload across general work tasks (Chapter 3). In other words, we showed support for objectively measuring mental workload in real-world environments when certain factors are considered. This is because after creating an ecologically valid environment in which participants were subjected to different and personalised general tasks and verbal interruptions, and mostly unrestricted coffee drinking, we found that fNIRS was still sensitive to mental workload levels to a large degree. By identifying how the sensitivity could have further been improved through carefully considering or expanding the measurement region and how relating findings to theoretical backgrounds can help to make sense of the data, we speculated that fNIRS would be effective for measuring real-world mental workload levels.

Tracking mental workload data in the real world is developing into a machine learning problem, comparably to how physical activity made the transition in the 2000s after the proof of concept stage and laboratory tracking stage [236]. Machine learning research for the measurement of mental workload is largely still at a laboratory stage, where research is working towards the accurate classification of levels using controlled data sets [27]. Research in this sense is increasing in quantity and accuracy [198; 27; 3; 163]. Our findings from Chapter 3 could be used to guide machine learning research when it progresses to real-world measurements.

This is because we identified factors that may affect the transition between laboratory environments to more uncontrolled environments. In this respect, we illustrated how real-world measurements are comprised of tasks of varying natures that may not always provoke responses in overlapping regions of the brain. In addition, when evaluating the accuracy of the classifications in real-world environments, we outlined how theoretical frameworks should be taken into account for the interpretation of data; unlike physical activity data that can quantify e.g. 500 steps, mental workload is not a tangible concept [66]. Indeed, the Multiple Resource Model [231] was an important tool for making sense of the data regarding the effect that interruptions had on mental workload levels in our study. Taken together, this chapter highlighted that 1) sensitivity will be one of the key challenges for automatically detecting mental workload changes from physiological data in the future, and 2) interpreting data in relation to mental workload models may be valuable for determining the accuracy of real-world mental workload measures.

8.2.2 Contribution 2 - The Mental Workload Cycle

The second contribution is the development of the Mental Workload Cycle, a new model that describes how people should aim to fluctuate between different mental workload levels in certain patterns, as sustaining any level for too long results in negative consequences, and each level serves a different and important purpose. This contribution emerged from approaching mental workload in a unique way in terms of taking a longitudinal, holistic, and person centred perspective. This is in contrast to the controlled, short-term, and isolated task based perspective that is widely adopted in research, and thus led to different types of insights than what has previously been seen in the literature. This improved understanding of mental workload progresses understanding for personal informatics (and future pBCI devices) and human factors, described below.

For Personal Informatics

The personal informatics field [82; 145; 194; 196] has not yet identified what cognitive data could be valuable to track and set goals for longitudinally in daily life in order to make life improvements. The Cycle findings, however, suggest that people could aim to follow a Mental Workload Cycle that is personalised for them in terms of levels and fluctuations; this could result in improvements to wellbeing and daily performance on tasks by preventing the negative consequences from sustaining any level for too long and by benefitting from the positive implications that each level can provide. Therefore, it seems as though finding the right balance of 'mental exercise' could be the main goal in terms of mental workload; similarly to physical activity, we should aim towards incorporating periods of rest, intense, and moderate activity into our lives.

Conceptually this way, performance on certain daily tasks could be optimised if people aim to structure their mental workload fluctuations in accordance with these tasks. For example, if someone wanted to optimise their performance on a high mental workload gaming session, they could self-moderate by limiting the amount of high mental workload they have before they game in order to make sure they do not get fatigued from spending too much time at that level. Instead they could structure their time before the gaming session to include a period of low mental workload to serve as a period of mental recovery and preparation before performing the high mental workload task. Thus, the identification of the Mental Workload Cycle has contributed to our understanding of how mental workload impacts our daily lives and what goals we should set.

These findings in terms of how mental workload impacts our lives are intuitive to understand. The link between mental workload and wellbeing factors, such as stress, is not such groundbreaking news. However, with the exception of sleep implications and fatigue [101], research has not previously identified the effect of daily mental workload levels on other factors relating to our lives. We have qualitatively and quantitatively identified several implications of mental workload. Companies concerned with developing pBCI technology, which is rapidly developing and increasingly filtering onto the consumer market [14; 16], could be guided by these early findings for the development of future pBCI devices in terms of what data would be valuable for users to reflect on. Our data also indicated that people frequently do not fluctuate in the patterns they prefer. Therefore, features of future devices should help users to modify their habits in order to implement their Cycles, and also create strategies that mitigate the negative effects for people who are not able to facilitate their Cycles.

Personal informatics tools that are already in existence could help to inform the initial design of a mental workload personal cognitive informatics tool. This is because we can draw on features of current psychological tracking tools that seem to reflect aspects of the Mental Workload Cycle.

In these terms, it is known that including moderate and vigorous physical exercise into our lives brings a myriad of health benefits[1], and also resting our bodies is needed for repairing and recovering. Physical activity tracker company Fitbit[2] provide users with a Daily Readiness Score, which combines recent physical activity data with sleep scores and heart rate variability in order to assess whether the user is ready to participate in physical activity or should prioritise rest instead. Similarly, another physical activity tracker company, Whoop[3], provides a daily recovery score by assessing heart rate variability, resting heart rate data, respiratory rate, sleep performance, blood oximetry, and skin temperature to determine the level of physical activity intensity the user could cope with (moderate or intense), and indeed whether they would benefit from resting instead.

In both of the above examples, the higher the readiness or recovery score, the more physically prepared the user is for taking on a larger and more vigorous physical load. This can be translated to the Mental Workload Cycle, as an individual should be more prepared to take on a higher mental workload load when they have achieved their proper Cycle fluctuations. In contrast, when, for example, an individual has over-sustained a high mental workload level, they might benefit from incorporating more low mental workload, or 'rest' into their day in order to recover. Thus, the idea of a daily recovery or readiness score might be useful for the design of a mental workload tracking device. This could be calculated from the

[1] https://www.nhs.uk/live-well/exercise/exercise-guidelines/physical-activity-guidelines-for-adults-aged-19-to-64/
[2] https://www.fitbit.com/global/us/technology/daily-readiness-score
[3] https://www.whoop.com/experience/recovery/

previous day's data (and perhaps sleep data), and could advise people about the type of 'mental workload day' that they are ready for, suggesting a Cycle based around that. The examples also reflect the notion that 'more' physical exercise is not always better, and we should not always aim for increased load. In terms of mental workload, our findings reflect this notion, as balance is key in order to benefit many big-picture aspects of ourselves (e.g. improved heart health for physical activity or reduced stress for mental workload) as opposed to the instant gratification from an immediate outcome (e.g. achieving a personal best for physical activity or completing a difficult task for mental workload).

Fitbit and Whoop also offer a sleep tracking function. Fitbit explains that humans go through several sleep cycles each night that each last for about 90 minutes, first consisting of light sleep, then deep sleep, and then Rapid Eye Movement (REM) sleep[4]. It is also explained that each sleep stage serves an important purpose, such as light sleep is important for physical and mental recovery, deep sleep is important for learning and memory, and REM sleep is important for mood regulation. Whoop mentions that getting enough sleep is important for cognitive functioning, stronger immune system, and overall metabolic health[5]. The characteristics of the Mental Workload Cycle share similar properties to these aspects of sleep cycles in terms of fluctuating between low, medium, and high levels in certain patterns, as each level serves an important purpose, and achieving a good Cycle appears to positively benefit several aspects of wellbeing and performance. Based on the sleep data, Fitbit provides a 'sleep score' that informs users about how well they slept, and information is offered for improving their scores and trends can be shown over time. These features could also be used to guide the design of a mental workload tracking device in terms

[4] https://help.fitbit.com/articles/en$_{U}S/Help_{a}rticle/2163.htm$

[5] https://www.whoop.com/thelocker/everything-to-know-about-sleep/

of informing users as to whether they have achieved their Cycles, advice for improving their Cycle adherence if they are not meeting their targets, and their adherence trends over time. Furthermore, Fitbit offers a 'Smart Wake' feature, where it wakes up the user at the optimal time in their sleep cycle. Future mental workload tracking devices could translate this feature to alerting users at the optimal time to take a break from their high mental workload level. Thus, taken together, there are several relatable aspects of currently available personal informatics tracking tools that future mental workload personal cognitive informatics tools could initially be guided by.

For Human Factors

These findings could also have a large impact on mental workload research in the human factors field. This is because we have illustrated how important it is to consider mental workload as a whole rather than focusing narrowly on mental workload limits within specific tasks. The novelty of the longitudinal approach to mental workload that was adopted in this thesis should be emphasised; mental workload literature traditionally considers mental workload from an input/output perspective [204]. Considering low, medium, and high levels throughout the day, and considering them both outside and inside the workplace, seems important for understanding performance on isolated work tasks. If mental workload levels outside of the workplace are not considered, we cannot understand the needs of people at work, and why they approach tasks in the way that they do. Our results suggest that sustained periods of (e.g.) high mental workload outside of work (perhaps with coordinating family life, as was qualitatively reported), can lead to fatigue more quickly during a work task that requires high mental workload, as the overall time spent at that level was greater than the time just spent on the work task. In essence, this could explain why performance on the same task completed on different days can vary.

Also, we identified that certain factors can interfere with people's ability to function at certain mental workload levels, which is also likely to contribute to varied task performance. This could mean that even if tasks are designed to be within a manageable mental workload level, factors taking place outside of the workplace might mean that people sometimes find them unmanageable. These findings echo work in related fields looking at the relationships between demands at home and safety at work [128], and visa versa [75; 74], as well as the participants who qualitatively described how carefully managed rest, both at work during breaks, and in the evenings, was needed to manage work. Thus, considering mental workload levels in between the extremes of overload and underload [240; 203] and taking a broader life perspective of mental workload could be essential for deepening understanding of factors that contribute to the 'redline' of these states.

In terms of implications for industry in relation to this, participants discussed their strategies within work tasks (at work and at home) that aimed to break up their high mental workload tasks with low mental workload tasks. A key factor in research focusing on the Future of Work, is to decompose work such that different sized tasks can be handled conveniently as e.g. microtasks [213]. This seems beneficial for people that have agency in their work, such as many office workers or self employed people. There are many jobs, however, such as air traffic controllers and train signal operators that have been carefully organised into shift patterns with predetermined breaks, where it is the responsibility of the employees to manage their ability to sustain mental workload accordingly. If mental workload is considered from a more holistic and person-centred perspective, tracking mental workload outside of work may enable workplaces and employees to manage mental workload more effectively during work tasks, especially as our findings suggested that people were not always able to implement their ideal mental workload strategies. Therefore, tracking mental workload in everyday life could be useful for improving safety and performance in safety-critical jobs.

8.2.3 Contributions 3 and 4 - Practical Guidance for pBCI Development

Once we know how mental workload data is valuable to track as personal informatics, and once the technology is developed enough to facilitate physiological mental workload tracking, how do we make interactions with mental workload personal informatics tools meaningful and safe?

Contribution 3 is therefore design recommendations for mental workload data. This was in terms of types of display, the use of customisable metaphors for visualisations, colours, shapes, and terminology. In particular, the importance of considering the differences between how mental workload cognitive activity is represented compared to how physical activity is represented was identified. Physical activity trackers use visualisations to represent more activity as positive, such as encouraging users to close their daily rings by participating in more physical activity. However, this type of visualisation does not apply to mental workload data, as sustaining levels for too long appears to have unhealthy consequences. The goals for mental workload compared to physical activity are therefore very different, and visualisations should reflect the importance of including each mental workload level alongside its optimal sustainment length. Colours may be a particularly meaningful way of communicating this to users. There is a propensity for traffic light colours to be used to communicate different levels of a measurable concept, yet this may misrepresent mental workload levels as either positive or negative, when they can be both; instead, changing vibrancies may be an effective way of communicating important aspects of mental workload data, such as sustainment lengths and fluctuations.

Thus, these findings should specifically implicate design aspects of future mental workload pBCI devices. Indeed, without effectively communicating personal data to users, the most impressive technology and collected data could be made meaningless.

Contribution 4 is the empirical identification of user ethical perceptions and concerns surrounding the development of pBCI devices. Regarding this, huge ethical challenges have been identified for the development of pBCI devices [139; 73; 228; 125; 216], but no research had looked at this empirically and from a users' perspective. Identifying user concerns relating to data misinterpretation and perceptions regarding safety-critical workers, we have added these considerations 'to the pile' of ethical challenges for the development of neurotechnology and further validated concerns relating to privacy, personal identity, and data validity.

These insights are therefore useful for the neuroethics field, which is driven by theoretical modifications of existing ethical frameworks and is lacking in real-world studies. Our research in this area could implicate further research and the revision of guidelines for the development of pBCI devices in this field. These findings 'should' also be useful for companies developing pBCI devices, but currently available devices have thus far found ways to avoid being held ethically accountable [228]. In order for ethics research regarding consumer neurotechnology to make a real difference to industry, there is a need for new regulations to be written into law so that users can be properly protected from ethical unintended consequences.

8.3 Future Research

As mentioned, research is making huge progress in measuring cognitive activity using physiological measures in real-world environments [158; 141]. Indeed, this includes mental workload measures, which are making promising progress [240]. Figure 8.1 compares the progress of cognitive activity trackers to how physical activity trackers progressed into what we see today [236]. What is interesting, is the division between the development of the technology and the understanding of the users. In this respect, whilst it is undoubtedly easier to develop meaningful uses for the technology once it can be studied in the wild, it seems sensible to consider that a deep understanding of how the data could be used to make meaningful life improvements could be developed in tandem with the development of the technology. Thus, when longitudinal tracking of real-world cognitive activity becomes available, it may be possible to provide users with meaningful personal cognitive informatics tools at a much greater speed.

The research in this thesis has contributed an understanding of the *type* of cognitive data that could be valuable as personal informatics, in terms of the Mental Workload Cycle. The next steps are to develop the Cycle on a more granular level in order to progress towards being able to practically implement the use of mental workload as personal informatics.

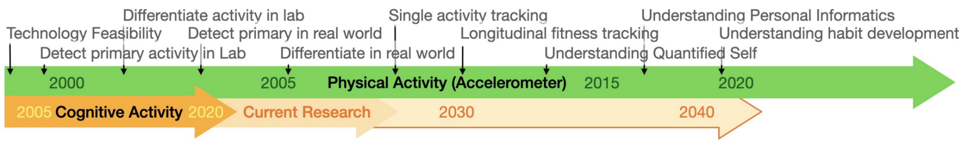


Figure 8.1: Where cognitive activity tracking is in relation to the development of physical activity tracking [236].

8.3.1 The Next Study

If this project was to be followed up with a post-doctoral fellowship, the next proposed study would aim to improve understanding of how people try to manage their mental workload levels in their daily lives. Our findings so far have related to participants' preferences for transitioning between the different mental workload levels, and we have started to comprehend the transitions that participants behaviourally made on a broad level. By building a more sensitive awareness of the transitions that people behaviourally make (e.g. breaking up high mental workload periods with quick low mental workload breaks, which would not have been captured in the quantitative data, but was described as a preferable pattern in the qualitative data), and supporting that data with qualitative insights specifically about management strategies in relation to the behavioural transitions, research would progress further towards understanding how to actually design healthy and efficient Mental Workload Cycles for individuals.

This proposed study could use another longitudinal mixed methods approach. Participants could be recruited for 3-5 days. Experience sampling would be used, where each day participants would be required to input their 'current' mental workload level. This means that every time their mental workload level changed, even for a short period of time, this data would be captured. At the end of each day, the participant would be interviewed by the researcher about their 'mental workload day' in terms of how they managed their mental workload, the effect it had on their performances and wellbeing, how it could have been better managed, and how they are feeling at the end of the day. Thus, the data could be analysed to start inferring links between more detailed mental workload patterns and its associated outcomes.

8.3.2 The Two to Five Year Plan

The findings from the next proposed study described above could provide a more detailed view of the type of lengths that people should sustain each mental workload level for before transitioning to another level, and how this might vary between individuals. What the findings will not show, however, is the effect on the individual after each block of mental workload (e.g. a period of predominantly low mental workload) as the research would still consider the data in terms of a day, which likely involves several fluctuations between levels.

With more time over the longer term, perhaps a period of two-five years, research could concentrate on further narrowing down on designing a functional Mental Workload Cycle that can be trialled for use as a personal informatics tool. This would require investigating the effects of different blocks of mental workload in terms of the effects it has performance and wellbeing and then combining the parts to develop a full Cycle that is ready for trial. The initial research could be inspired by Mark et al [154] who used longitudinal ethnographic observations and interviews on information workers to determine how work tasks are fragmented, the causes of the fragmentation, and the effect of the fragmentation on task completion. Their method involved detailed descriptions of each participant's activities (such as opening a notebook) for three and and half days. Our research could aim to collect this level of detail, but in terms of switching between mental workload levels instead of switching between tasks/interruptions. However, as our previous research has already informed us about the optimal patterns of fluctuations, and by this point should have provided an idea of how long each level should be sustained for, we could implement a level of mental workload manipulation. In this respect, participants could be asked to maintain certain levels of mental workload for a certain period of time before the effect of that, and its cumulative effect with other mental workload levels, is evaluated.

Thus, the study design would involve the researcher to be present whilst the participant is at work. The researcher would then ask the participant to stay at e.g. a high mental workload level for a designated period of time (e.g. 40 minutes) before dropping to a low mental workload level for e.g. 5 minutes. Based on how the participant perceived the starting block in terms of performance and wellbeing (e.g. fatigue levels), subsequent blocks could be adapted in length or adapted depending on type of task (e.g. enjoyable vs not enjoyable); the researcher could encourage participants to direct the block lengths according to how they feel, but stick to the patterns of fluctuations that have previously been identified as optimal. Questionnaires or interviews after each block and at the end of each day could evaluate the effect that different blocks, and the combination of blocks deliberately dictated by the participant, had on efficiency, perceptions, and wellbeing. A gradual, detailed understanding of how mental workload is optimally managed in everyday life could be developed from this.

The study could also be expanded to investigate mental workload outside of work, which may benefit from managing mental workload levels differently to during work. This would be especially interesting to investigate from a stress point of view. This is because, theoretically, mental workload and stress are related concepts with overlapping characteristics in terms of perceived response capability [8]. The relationship between mental workload and stress that has been identified in this thesis could somewhat reflect the idea that high mental workload tasks are often stressful, for example having an element of time pressure. It would therefore be interesting to investigate the effects of high mental workload, low stress tasks (which might be seen more in a home environment) and the effects of high mental workload, high stress tasks, in terms of general stress in life. Thus, this could impact the insights provided for personal informatics, such that more high mental workload tasks might be beneficial to cognitive health if they are not associated with stress.

After developing an understanding of how we can design these Mental Workload Cycles, a further longitudinal study could design a Cycle for each participant based on individual characteristics. Participants could then aim to follow their Cycle for a week, before being interviewed once more in depth about the effects of their mental workload management.

8.3.3 The Ten to Twenty Year Outlook

In the long-term, there should be a clear understanding of how mental workload could be used for personal informatics, and tools to facilitate this are likely to be readily available. As we saw with physical activity tracking (Figure 8.1), the next step is to understand habit formation. In this respect, changing behaviour is a complex process [136], but once action to implement change has been taken, developing habits are important for ensuring behaviour change has long-term effects [210]. Whilst theories of habit formation can vary, all consider habits as repetitions of behaviour in a particular context that results in a transition from goal-directed behaviour to automatic habitual behaviour through repeated learning [40]. Research by Stawarz et al [210] has shown, however, that apps designed specifically to support habit change tend to focus on self-tracking, which is an important aspect for behaviour change but does not support habit formation. Thus, the researchers outlined guidance for HCI researchers designing apps to better support habit formation.

Therefore, the long-term goal for research into mental workload for use as personal informatics would be to support habit formation in order for mental workload management behaviours to become less deliberate and more embedded into everyday life. A starting point for integrating habit formation support into mental workload personal informatics tools would be to design features based on the recommendations outlined by Stawarz et al [210]. In this respect, trigger events could be incorporated, which help users to form implementation intentions by explicitly stating their behavioural action plan in order to reinforce the association between an event and a behaviour. In addition, the tool could explore the use of reminders before the trigger event in order to reinforce the association but still encourage the user to independently make the association when the event occurs [210].

For mental workload, this could, perhaps, relate to users forming associations between mental workload fluctuation patterns and level sustainment lengths. For example, "Don't forget to take a low mental workload level break after your high mental workload period." This way, it could be hypothesised that users would be supported in having agency over their mental workload management, becoming more immersed and invested in their behaviour change. This could perhaps lead to increased feelings of fulfillment in the form of intrinsic rewards, an area that is not yet well understood [210], but has recently been shown to play a role in habit formation for healthy eating behaviours [68]. In any case, planning to support the development of habits for mental workload management can help us to envision a future of wide-spread mental workload tracking that supports long-term healthy and efficient outcomes, perhaps in a similar way to how we currently see physical activity habits often being embedded into people's lives.

Chapter 9

Thesis Conclusions

As life is becoming increasingly characterised by its cognitive complexity, it seems sensible to assume that the management of our cognitive activity in daily life is becoming increasingly important. It appears that this need is being recognised, and companies are beginning to produce consumer pBCI devices en masse. This thesis was motivated by aiming to understand mental workload tracking in daily life and how life improvements could be made through using this data as a form of personal informatics. This could be useful for the development of future pBCI devices as they progress towards longitudinal tracking. We adopted a longitudinal and person-centred perspective of mental workload, which is in contrast to typical approaches that consider mental workload as an input/output model [204]. The overarching research question for this thesis was:

- How and why should we track mental workload in everyday life?

Two empirical studies were conducted that addressed four sub-questions, each of which will be outlined in turn below.

The first sub-question aimed to measure mental workload for tasks relevant to daily life, in terms of the ability of physiological measures to differentiate between different mental workload levels and the challenges of doing so in real-world environments. In this respect, the first sub-question was:

(a) Can we physiologically track mental workload levels in general work tasks, and what are the practical concerns of doing that?

The first empirical study addressed this, outlined in Chapter 3. The study was a laboratory study that included aspects representative of real-world work environments. fNIRS was used as the physiological measure due to its promise for real-world measurements [189; 115]. 19 participants completed naturalistic reading and writing tasks at different levels of difficulty; the tasks were personalised to each participant, verbal interruptions were incorporated, and coffee drinking was unrestricted. Results found that fNIRS was able to detect differences in mental workload levels for the reading tasks but not the writing tasks. This finding highlighted the sensitivity challenges that real-world mental workload measurements might face. Further, mental workload models were instrumental for understanding the impact that verbal interruptions had on mental workload levels. Taken together, these findings could impact future machine learning research and made the following contribution:

- Identified challenges for measuring mental workload across general work tasks.

The aim of the second sub-question was to adopt a novel longitudinal approach to mental workload in order to investigate how the data could be used for personal informatics, whilst also developing a deeper understanding of the mental workload concept. Thus, the second sub-question was:

(b) Can a longitudinal and holistic approach to mental workload improve understanding of how mental workload could be valuable as a form of personal informatics, and mental workload as a concept itself?

The second empirical study addressed this sub-question. The study has a quantitative and qualitative phase and was outlined in Chapter 4. The quantitative phase involved 19 participants that subjectively tracked their mental workload levels at regular intervals from Monday-Friday; phone activity was recorded and questionnaire data relating to wellbeing aspects was collected each evening. The qualitative phase involved the same participants and interviewed participants in depth about their personal experiences of mental workload.

The qualitative phase was analysed first; the large volume of data from the quantitative phase could then be guided in respect to the insights gained from the qualitative research, which was beneficial for determining the direction of the analysis. Data from the qualitative phase identified how fluctuating between mental workload levels in certain patterns seems important for aspects of wellbeing, daily performances, and positive perceptions; this was termed the Mental Workload Cycle (Figure 9.1). Data from the quantitative phase then indicated that actual daily life behaviours in terms of fluctuations frequently differed to qualitative preferences, as well as further identifying the negative impacts that mental workload can have on aspects of our wellbeing. (Chapter 5.) Thus, these findings demonstrate how mental workload could be used as a form of personal cognitive informatics.

The traditional understanding of mental workload was also progressed through identifying the Mental Workload Cycle. This was because by understanding person-centred perceptions, life approaches, and life implications, a better understanding of the factors that contribute to the states of overload and underload were developed.

Taken together, these findings should progress research in the personal informatics and human factors fields, and could be used to progress the design of future pBCI devices. Specifically, the following contribution was made:

- Developed a new model of mental workload, specifically the Mental Workload Cycle, which improves our understanding of mental workload and how that data can be used as personal cognitive informatics.

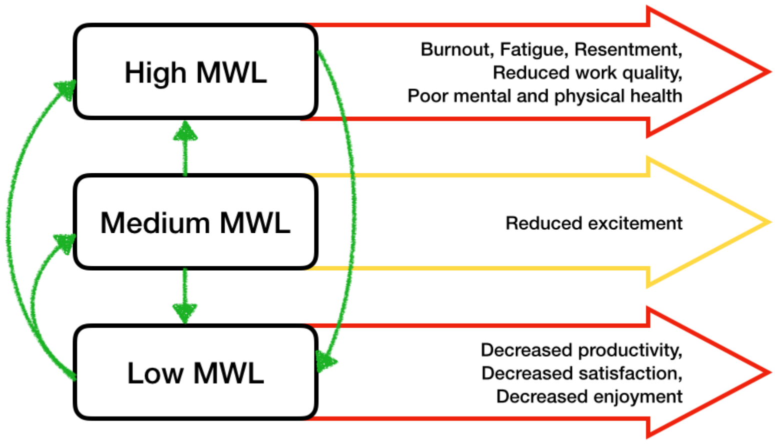


Figure 9.1: The Mental Workload Cycle, which was the main contribution from this thesis.

The qualitative phase also explored user perceptions relating to the interaction between users and objective mental workload tracking devices. This included researching data communication aspects, aiming to facilitate meaningful interactions between users and the technology. Therefore the third sub-question regarded:

(c) How can objective mental workload tracking data be meaningfully communicated to users?

The findings were outlined in Chapter 6 and design recommendations were developed from this, which companies could use to effectively communicate mental workload data to users. Therefore, the following contribution was made:

- Produced design recommendations for mental workload data.

The other interaction aspect that was investigated in Study 3 was ethical perceptions, which aimed to facilitate safe interactions with the technology. The final sub-question was therefore:

(d) What should be ethically considered when developing mental workload pBCI devices, or neurotechnology in general?

Ethical concerns from potential users of future pBCI devices were identified, outlined in Chapter 7; these findings should be generalisable beyond specific mental workload technologies, and could affect guidelines produced in the neuroethics field. The following contribution from this was made:

- Explicated ethical concerns for the development of pBCIs.

9.1 Final Remarks

Physical activity trackers for use as personal informatics have reached a stage where they can be considered ubiquitous. With rapid progress in brain imaging technology, the next logical, realistic, and exciting stage is to consider longitudinal cognitive data tracking for personal *cognitive* informatics. Perhaps uncoincidentally, this progress in technology development has coincided with an increasing need to better manage the cognitive activity in our lives in order to meet the growing cognitive demands of work life and home life, whilst maintaining a high level of performance and well-being. This thesis has considered mental workload from a longitudinal, holistic, and person-centred perspective, and this novel approach has been fundamental for understanding how mental workload can be valuable as a form of personal informatics. In this respect, the main contribution from this thesis is the Mental Workload Cycle, which outlines how we should aim to fluctuate between levels in certain patterns in order to optimise aspects of our performances, perceptions, and wellbeing. These findings are promising for the future and hopefully research will continue to build upon the Cycle in order to facilitate a transition towards wide-spread tracking of mental workload as a form of personal cognitive informatics.

Bibliography

[1] Jans Aasman, Gijsbertus Mulder, and Lambertus JM Mulder. 1987. Operator effort and the measurement of heart-rate variability. *Human factors* 29, 2 (1987), 161–170.

[2] Christopher M Aasted, Meryem A Yücel, Robert J Cooper, Jay Dubb, Daisuke Tsuzuki, Lino Becerra, Mike P Petkov, David Borsook, Ippeita Dan, and David A Boas. 2015. Anatomical guidance for functional near-infrared spectroscopy: AtlasViewer tutorial. *Neurophotonics* 2, 2 (2015), 020801.

[3] Daniel Afergan, Evan M Peck, Erin T Solovey, Andrew Jenkins, Samuel W Hincks, Eli T Brown, Remco Chang, and Robert JK Jacob. 2014. Dynamic difficulty using brain metrics of workload. In *Proc. SIGCHI*. ACM, 3797–3806.

[4] Ulf Ahlstrom and Ferne J Friedman-Berg. 2006. Using eye movement activity as a correlate of cognitive workload. *International journal of industrial ergonomics* 36, 7 (2006), 623–636.

[5] Minkyu Ahn, Mijin Lee, Jinyoung Choi, and Sung Chan Jun. 2014. A review of brain-computer interface games and an opinion survey from researchers, developers and users. *Sensors* 14, 8 (2014), 14601–14633.

[6] Icek Ajzen, Thomas C Brown, and Franklin Carvajal. 2004. Explaining the discrepancy between intentions and actions: The case of hypothetical bias in contingent valuation. *Personality and social psychology bulletin* 30, 9 (2004), 1108–1121.

[7] Icek Ajzen and James Sexton. 1999. Depth of processing, belief congruence, and attitude-behavior correspondence. *Dual-process theories in social psychology* (1999), 117–138.

[8] Norah H Alsuraykh, Max L Wilson, Paul Tennent, and Sarah Sharples. 2019. How stress and mental workload are connected. In *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 371–376.

[9] Lawrence L Ames and Edward J George. 1993. *Revision and verification of a seven-point workload estimate scale*. Technical Report. Air Force Test Center Edwards AFB CA.

[10] Pavlo Antonenko, Fred Paas, Roland Grabner, and Tamara Van Gog. 2010. Using electroencephalography to measure cognitive load. *Educational psychology review* 22, 4 (2010), 425–438.

[11] Elizabeth M Argyle, Adrian Marinescu, Max L Wilson, Glyn Lawson, and Sarah Sharples. 2021. Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments. *International Journal of Human-Computer Studies* 145 (2021), 102522.

[12] Pietro Aricò, Gianluca Borghini, Gianluca Di Flumeri, Alfredo Colosimo, Stefano Bonelli, Alessia Golfetti, Simone Pozzi, Jean-Paul Imbert, Géraud Granger, Raïlane Benhacene, et al. 2016. Adaptive automation triggered by EEG-based mental workload index: a pas-

sive brain-computer interface application in realistic air traffic control environment. *Frontiers in human neuroscience* 10 (2016), 539.

[13] P Aricò, G Borghini, G Di Flumeri, A Colosimo, S Pozzi, and F Babiloni. 2016. A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Progress in brain research* 228 (2016), 295–328.

[14] P Aricò, G Borghini, G Di Flumeri, N Sciaraffa, and F Babiloni. 2018. Passive BCI beyond the lab: current trends and future directions. *Physiological measurement* 39, 8 (2018), 08TR02.

[15] Pietro Arico, Gianluca Borghini, Gianluca Di Flumeri, Nicolina Sciaraffa, Alfredo Colosimo, and Fabio Babiloni. 2017. Passive BCI in operational environments: insights, recent advances, and future trends. *IEEE Transactions on Biomedical Engineering* 64, 7 (2017), 1431–1436.

[16] Pietro Aricò, Nicolina Sciaraffa, and Fabio Babiloni. 2020. Brain–computer interfaces: Toward a daily life employment. , 157 pages.

[17] Hasan Ayaz, Banu Onaral, Kurtulus Izzetoglu, Patricia A Shewokis, Ryan McKendrick, and Raja Parasuraman. 2013. Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. *Frontiers in human neuroscience* 7 (2013), 871.

[18] Hasan Ayaz, Patricia A Shewokis, Scott Bunce, Kurtulus Izzetoglu, Ben Willems, and Banu Onaral. 2012. Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 1 (2012), 36–47.

[19] Amid Ayobi, Tobias Sonne, Paul Marshall, and Anna L Cox. 2018. Flexible and mindful self-tracking: Design implications from paper bullet journals. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.

[20] Richard W Backs, H Ted Navidzadeh, and Xidong Xu. 2000. Cardiorespiratory indices of mental workload during simulated air traffic control. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 44. SAGE Publications Sage CA: Los Angeles, CA, 89–92.

[21] Alan Baddeley. 2012. Working memory: Theories, models, and controversies. *Annual review of psychology* 63 (2012), 1–29.

[22] Brian P Bailey and Shamsi T Iqbal. 2008. Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction (TOCHI)* 14, 4 (2008), 21.

[23] Joana B Balardin, Guilherme A Zimeo Morais, Rogério A Furucho, Lucas Trambaiolli, Patricia Vanzella, Claudinei Biazoli Jr, and João R Sato. 2017. Imaging brain function with functional near-infrared spectroscopy in unconstrained environments. *Frontiers in human neuroscience* 11 (2017), 258.

[24] Jakob E Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2627–2636.

[25] Luciane Baretta, Lêda Maria Braga Tomitch, Vanessa K Lim, and Karen E Waldie. 2012. Investigating reading comprehension through

EEG. *Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies* 63 (2012), 69–99.

[26] Lyn Bartram, Abhisekh Patra, and Maureen Stone. 2017. Affective color in visualization. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 1364–1374.

[27] Johann Benerradi, Horia A. Maior, Adrian Marinescu, Jeremie Clos, and Max L. Wilson. 2019. Exploring machine learning approaches for classifying mental workload using fNIRS data from HCI tasks. In *Proceedings of the Halfway to the Future Symposium 2019*. 1–11.

[28] Chris Berka, Daniel J Levendowski, Milenko M Cvetinovic, Miroslav M Petrovic, Gene Davis, Michelle N Lumicao, Vladimir T Zivkovic, Miodrag V Popovic, and Richard Olmstead. 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 (2004), 151–170.

[29] Matt T Bianchi, Kathryn L Williams, Scott Mckinney, and Jeffrey M Ellenbogen. 2013. The subjective–objective mismatch in sleep perception among those with insomnia and sleep apnea. *Journal of sleep research* 22, 5 (2013), 557–568.

[30] Benjamin Blankertz, Laura Acqualagna, Sven Dähne, Stefan Haufe, Matthias Schultze-Kraft, Irene Sturm, Marija Uščumlic, Markus A Wenzel, Gabriel Curio, and Klaus-Robert Müller. 2016. The Berlin brain-computer interface: progress beyond communication and control. *Frontiers in neuroscience* 10 (2016), 530.

[31] Gianluca Borghini, Pietro Aricò, Laura Astolfi, Jlenia Toppi, Febo Cincotti, Donatella Mattia, Patrizia Cherubino, Giovanni Vecchiato,

Anton Giulio Maglione, Ilenia Graziani, et al. 2013. Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 6619–6622.

[32] Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews* 44 (2014), 58–75.

[33] Brian Bourke. 2014. Positionality: Reflecting on the research process. *The qualitative report* 19, 33 (2014), 1–9.

[34] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. (2012).

[35] Virginia Braun and Victoria Clarke. 2013. *Successful qualitative research: A practical guide for beginners.* sage.

[36] SD Brennan. 1992. An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder. *Portsmouth: DRA Maritime Command and Control Division. DRA Technical Memorandum (CAD5)* 92017 (1992).

[37] Jeffrey B Brookings, Glenn F Wilson, and Carolyne R Swain. 1996. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology* 42, 3 (1996), 361–377.

[38] Anne-Marie Brouwer, Maarten A Hogervorst, Jan BF Van Erp, Tobias Heffelaar, Patrick H Zimmerman, and Robert Oostenveld. 2012. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering* 9, 4 (2012), 045008.

[39] Brad Cain. 2007. *A review of the mental workload literature.* Technical Report. DTIC Document.

[40] Lucas Carden and Wendy Wood. 2018. Habit formation and change. *Current opinion in behavioral sciences* 20 (2018), 117–122.

[41] Giulia Cartocci, Myriam Caratù, Enrica Modica, Anton Giulio Maglione, Dario Rossi, Patrizia Cherubino, and Fabio Babiloni. 2017. Electroencephalographic, heart rate, and galvanic skin response assessment for an advertising perception study: application to anti-smoking public service announcements. *JoVE (Journal of Visualized Experiments)* 126 (2017), e55872.

[42] John G Casali and Walter W Wierwille. 1984. On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics* 27, 10 (1984), 1033–1050.

[43] Antonio Cepeda-Benito, David H Gleaves, Tara L Williams, and Stephen A Erath. 2000. The development and validation of the state and trait food-cravings questionnaires. *Behavior therapy* 31, 1 (2000), 151–173.

[44] Justin Chan, Sarah Power, and Tom Chau. 2012. Investigating the need for modelling temporal dependencies in a brain-computer interface with real-time feedback based on near infrared spectra. *Journal of Near Infrared Spectroscopy* 20, 1 (2012), 107–116.

[45] Rebecca L Charles and Jim Nixon. 2019. Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics* 74 (2019), 221–232.

[46] Patrizia Cherubino, Arianna Trettel, Giulia Cartocci, Dario Rossi,

Enrica Modica, Anton Giulio Maglione, Marco Mancini, Gianluca di Flumeri, and Fabio Babiloni. 2016. Neuroelectrical indexes for the study of the efficacy of TV advertising stimuli. In *Selected issues in experimental economics*. Springer, 355–371.

[47] Gladys Lai Cheng Chia, Angelika Anderson, and Louise Anne McLean. 2019. Behavior change techniques incorporated in fitness trackers: Content analysis. *JMIR mHealth and uHealth* 7, 7 (2019), e12768.

[48] Eun Kyoung Choe, Saeed Abdullah, Mashfiqui Rabbi, Edison Thomaz, Daniel A Epstein, Felicia Cordeiro, Matthew Kay, Gregory D Abowd, Tanzeem Choudhury, James Fogarty, et al. 2017. Semi-automated tracking: a balanced approach for self-monitoring applications. *IEEE Pervasive Computing* 16, 1 (2017), 74–84.

[49] Eun Kyoung Choe, Bongshin Lee, et al. 2015. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE computer graphics and applications* 35, 4 (2015), 28–37.

[50] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 173–182.

[51] Avishek Choudhury and Onur Asan. 2021. Impact of using wearable devices on psychological Distress: Analysis of the health information national Trends survey. *International journal of medical informatics* 156 (2021), 104612.

[52] Febo Cincotti, Donatella Mattia, Fabio Aloise, Simona Bufalari, Gerwin Schalk, Giuseppe Oriolo, Andrea Cherubini, Maria Grazia Mar-

ciani, and Fabio Babiloni. 2008. Non-invasive brain–computer interface system: towards its application as assistive technology. *Brain research bulletin* 75, 6 (2008), 796–803.

[53] Luigina Ciolfi and Eleanor Lockley. 2018. From Work to Life and back again: Examining the digitally-mediated work/life practices of a group of knowledge workers. *Computer Supported Cooperative Work (CSCW)* 27, 3 (2018), 803–839.

[54] Tom Clarke and Alan Costall. 2008. The emotional connotations of color: A qualitative investigation. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 33, 5 (2008), 406–410.

[55] Sheldon Cohen, Tom Kamarck, Robin Mermelstein, et al. 1994. Perceived stress scale. *Measuring stress: A guide for health and social scientists* 10, 2 (1994), 1–2.

[56] Margaretta Colangelo. 2020. How AI Is Advancing NeuroTech. *Forbes.com* (Feb 2020). https://www.forbes.com/sites/cognitiveworld/2020/02/12/how-ai-is-advancing-neurotech/

[57] C Collet, E Salvia, and C Petit-Boulanger. 2014. Measuring workload with electrodermal activity during common braking actions. *Ergonomics* 57, 6 (2014), 886–896.

[58] Emily IM Collins, Anna L Cox, Jon Bird, and Cassie Cornish-Tresstail. 2014. Barriers to engagement with a personal informatics productivity tool. In *Proceedings of the 26th Australian Computer-*

Human interaction Conference on Designing futures: The future of design. 370–379.

[59] Sunny Consolvo, David W McDonald, and James A Landay. 2009. Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 405–414.

[60] George E Cooper and Robert P Harper. 1969. *The use of pilot rating in the evaluation of aircraft handling qualities*. National Aeronautics and Space Administration.

[61] Shelly L Curran, Michael A Andrykowski, and Jamie L Studts. 1995. Short form of the profile of mood states (POMS-SF): psychometric information. *Psychological assessment* 7, 1 (1995), 80.

[62] Edward Cutrell and Desney Tan. 2008. BCI for passive input in HCI. In *Proceedings of CHI*, Vol. 8. Citeseer, 1–3.

[63] Ian Daly, Duncan Williams, Alexis Kirke, James Weaver, Asad Malik, Faustina Hwang, Eduardo Miranda, and Slawomir J Nasuto. 2016. Affective brain–computer music interfacing. *Journal of Neural Engineering* 13, 4 (2016), 046022.

[64] Prabu David, Jung-Hyun Kim, Jared S Brickman, Weina Ran, and Christine M Curtis. 2015. Mobile phone distraction while studying. *New media & society* 17, 10 (2015), 1661–1679.

[65] Michel De Rivecourt, MN Kuperus, WJ Post, and LJM Mulder. 2008. Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics* 51, 9 (2008), 1295–1319.

[66] Dick de Waard and Ben Lewis-Evans. 2014. Self-report scales alone cannot capture mental workload. *Cognition, technology & work* 16, 3 (2014), 303–305.

[67] Frédéric Dehais, Alban Dupres, Gianluca Di Flumeri, Kevin Verdiere, Gianluca Borghini, Fabio Babiloni, and Raphalle Roy. 2018. Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 544–549.

[68] Sally Di Maio, Jan Keller, Dominika Kwasnicka, Nina Knoll, Lena Sichert, and Lena Fleig. 2022. What helps to form a healthy nutrition habit? Daily associations of intrinsic reward, anticipated regret, and self-efficacy with automaticity. *Appetite* (2022), 106083.

[69] AC Dieler, SV Tupak, and AJ Fallgatter. 2012. Functional near-infrared spectroscopy for the assessment of speech related tasks. *Brain and Language* 121, 2 (2012), 90–109.

[70] Jillian Dorrian, Stuart D Baulk, and Drew Dawson. 2011. Work hours, workload, sleep and fatigue in Australian Rail Industry employees. *Applied ergonomics* 42, 2 (2011), 202–209.

[71] Duncan Douglas, Hoekstra Angel, and Wilcox Bethany. 2012. Digital devices, distraction, and student performance: Does in-class cell phone use reduce learning? *Astronomy education review* (2012).

[72] Gautier Durantin, J-F Gagnon, Sébastien Tremblay, and Frédéric Dehais. 2014. Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research* 259 (2014), 16–23.

[73] Margaret L Eaton and Judy Illes. 2007. Commercializing cognitive neurotechnology—the ethical terrain. *Nature biotechnology* 25, 4 (2007), 393–397.

[74] Achim Elfering, Christin Gerhardt, Diana Pereira, Anna Schenker, and Maria U Kottwitz. 2020. The Monday effect revisited: a diary and sleep actigraphy study. *Sleep and vigilance* 4, 2 (2020), 167–176.

[75] Achim Elfering, Simone Grebner, and Coralie Boillat. 2013. Busy at work and absent-minded at home: Mental workload, cognitive failure, and domestic falls. *Swiss Journal of Psychology* 72, 4 (2013), 219.

[76] Robert Elliott, Constance T Fischer, and David L Rennie. 1999. Evolving guidelines for publication of qualitative research studies in psychology and related fields. *British journal of clinical psychology* 38, 3 (1999), 215–229.

[77] David Elsweiler, Max L Wilson, and Brian Kirkegaard Lunn. 2011. Understanding casual-leisure information behaviour. In *New directions in information behaviour*. Emerald Group Publishing Limited.

[78] Daniel A Epstein. 2015. Personal informatics in everyday life. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 429–434.

[79] Daniel A Epstein, Clara Caldeira, Mayara Costa Figueiredo, Xi Lu, Lucas M Silva, Lucretia Williams, Jong Ho Lee, Qingyang Li, Simran Ahuja, Qiuer Chen, et al. 2020. Mapping and taking stock of the personal informatics literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–38.

[80] Daniel A Epstein, Jennifer H Kang, Laura R Pina, James Fogarty,

and Sean A Munson. 2016. Reconsidering the device in the drawer: lapses as a design opportunity in personal informatics. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. 829–840.

[81] Daniel A Epstein, Nicole B Lee, Jennifer H Kang, Elena Agapie, Jessica Schroeder, Laura R Pina, James Fogarty, Julie A Kientz, and Sean Munson. 2017. Examining menstrual tracking to inform the design of personal informatics tools. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6876–6888.

[82] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 731–742.

[83] Jordan Etkin. 2016. The hidden cost of personal quantification. *Journal of consumer research* 42, 6 (2016), 967–984.

[84] Chloe Fan, Jodi Forlizzi, and Anind K Dey. 2012. A spark of activity: exploring informative art as visualization for physical activity. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 81–84.

[85] Martha J Farah. 2012. Neuroethics: the ethical, legal, and societal impact of neuroscience. *Annual review of psychology* 63 (2012), 571–591.

[86] Tom Fawcett. 2015. Mining the quantified self: personal knowledge discovery as a challenge for data science. *Big data* 3, 4 (2015), 249–266.

[87] Sara B Festini, Christopher Hertzog, Ian M McDonough, and

Denise C Park. 2019. What makes us busy? Predictors of perceived busyness across the adult lifespan. *The Journal of general psychology* 146, 2 (2019), 111–133.

[88] Sara B Festini, Ian M McDonough, and Denise C Park. 2016. The busier the better: greater busyness is associated with better cognition. *Frontiers in aging neuroscience* 8 (2016), 98.

[89] Frank Anthony Fishburn, Megan E Norr, Andrei V Medvedev, and Chandan J Vaidya. 2014. Sensitivity of fNIRS to cognitive state and load. *Frontiers in human neuroscience* 8 (2014), 76.

[90] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.

[91] Gianluca Di Flumeri, Gianluca Borghini, Pietro Aricò, Alfredo Colosimo, Simone Pozzi, Stefano Bonelli, Alessia Golfetti, Wanzeng Kong, and Fabio Babiloni. 2015. On the use of cognitive neurometric indexes in aeronautic and air traffic management environments. In *International Workshop on Symbiotic Interaction*. Springer, 45–56.

[92] Sophie Forster and Nilli Lavie. 2007. High perceptual load makes everybody equal. *Psychological science* 18, 5 (2007), 377–381.

[93] Hannah J Foy and Peter Chapman. 2018. Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied ergonomics* 73 (2018), 90–99.

[94] Hannah J Foy, Patrick Runham, and Peter Chapman. 2016. Prefrontal cortex activation and young driver behaviour: a fNIRS study. *PLoS one* 11, 5 (2016), e0156512.

[95] Jon Froehlich, Tawanna Dillahunt, Predrag Klasnja, Jennifer Mankoff, Sunny Consolvo, Beverly Harrison, and James A Landay.

2009. UbiGreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the sigchi conference on human factors in computing systems*. 1043–1052.

[96] Jon Froehlich, Leah Findlater, Marilyn Ostergren, Solai Ramanathan, Josh Peterson, Inness Wragg, Eric Larson, Fabia Fu, Mazhengmin Bai, Shwetak Patel, et al. 2012. The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2367–2376.

[97] John DE Gabrieli, Russell A Poldrack, and John E Desmond. 1998. The role of left prefrontal cortex in language and memory. *Proceedings of the national Academy of Sciences* 95, 3 (1998), 906–913.

[98] Hermann Garden, Diana M Bowman, Sebastian Haesler, and David E Winickoff. 2016. Neurotechnology and society: Strengthening responsible innovation in brain science. *Neuron* 92, 3 (2016), 642–646.

[99] Edward Leigh Gibson. 2006. Emotional influences on food choice: sensory, physiological and psychological pathways. *Physiology & behavior* 89, 1 (2006), 53–61.

[100] James Giordano and Roland Benedikter. 2012. An early-and necessary-flight of the owl of Minerva: neuroscience, neurotechnology, human socio-cultural boundaries, and the importance of neuroethics. *J Evol Technol* 22, 1 (2012), 14–25.

[101] Namni Goel, Takashi Abe, Marcia E Braun, and David F Dinges. 2014. Cognitive workload and sleep restriction interact to influence sleep homeostatic responses. *Sleep* 37, 11 (2014), 1745–1756.

[102] Daniel Gopher and Emanuel Donchin. 1986. Workload: An examination of the concept. (1986).

[103] Mariel Grassmann, Elke Vlemincx, Andreas Von Leupoldt, Justin M Mittelstädt, and Omer Van den Bergh. 2016. Respiratory changes in response to cognitive load: a systematic review. *Neural plasticity* (2016).

[104] J Robert Grove and Harry Prapavessis. 1992. Preliminary evidence for the reliability and validity of an abbreviated profile of mood states. *International Journal of Sport Psychology* (1992).

[105] Jaime Guixeres, Enrique Bigné, Jose M Ausin Azofra, Mariano Alcaniz Raya, Adrian Colomer Granero, Felix Fuentes Hurtado, and Valery Naranjo Ornedo. 2017. Consumer neuroscience-based metrics predict recall, liking and viewing rates in online advertising. *Frontiers in psychology* 8 (2017), 1808.

[106] Charlotte C Gupta, Mitch J Duncan, Sally A Ferguson, Amanda Rebar, Madeline Sprajcer, Saman Khalesi, Lauren A Booker, Hannah Binks, and Grace E Vincent. 2021. The discrepancy between knowledge of sleep recommendations and the actual sleep behaviour of Australian adults. *Behavioral sleep medicine* 19, 6 (2021), 828–839.

[107] Eija Haapalainen, SeungJun Kim, Jodi F Forlizzi, and Anind K Dey. 2010. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, 301–310.

[108] GM Hancock, L Longo, MS Young, and PA Hancock. 2021. Mental workload. *Handbook of Human Factors and Ergonomics* (2021), 203–226.

[109] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[110] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload* (1988).

[111] Mariam Hassib, Mohamed Khamis, Stefan Schneegass, Ali Sahami Shirazi, and Florian Alt. 2016. Investigating user needs for bio-sensing and affective wearables. In *Proceedings of the 2016 Chi conference extended abstracts on human factors in computing systems*. 1415–1422.

[112] Stefan Haufe, Jeong-Woo Kim, Il-Hwa Kim, Andreas Sonnleitner, Michael Schrauf, Gabriel Curio, and Benjamin Blankertz. 2014. Electrophysiology-based detection of emergency braking intention in real-world driving. *Journal of neural engineering* 11, 5 (2014), 056011.

[113] Keith C Hendy, Kevin M Hamilton, and Lois N Landry. 1993. Measuring subjective workload: when is one scale better than many? *Human Factors* 35, 4 (1993), 579–601.

[114] Christian Herff, Dominic Heger, Ole Fortmann, Johannes Hennrich, Felix Putze, and Tanja Schultz. 2014. Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Frontiers in human neuroscience* 7 (2014), 935.

[115] Fabian Herold, Patrick Wiegel, Felix Scholkmann, and Notger G Müller. 2018. Applications of functional near-infrared spectroscopy

(fNIRS) neuroimaging in exercise–cognition science: a systematic, methodology-focused review. *Journal of clinical medicine* 7, 12 (2018), 466.

[116] Fabian Herold, Patrick Wiegel, Felix Scholkmann, Angelina Thiers, Dennis Hamacher, and Lutz Schega. 2017. Functional near-infrared spectroscopy in movement science: a systematic review on cortical activity in postural and walking tasks. *Neurophotonics* 4, 4 (2017), 041403.

[117] William J Horrey, Mary F Lesch, and A Garabet. 2009. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *Journal of safety research* 40, 1 (2009), 7–12.

[118] Alexander E Hramov, Vladimir A Maksimenko, and Alexander N Pisarchik. 2021. Physical principles of brain–computer interfaces and their applications for rehabilitation, robotics and control of human brain states. *Physics Reports* 918 (2021), 1–133.

[119] Bin-Wei Hsu, Mao-Jiun J Wang, Chi-Yuan Chen, and Fang Chen. 2015. Effective indices for monitoring mental workload while performing multiple tasks. *Perceptual and motor skills* 121, 1 (2015), 94–117.

[120] Dandan Huang, Melanie Tory, Bon Adriel Aseniero, Lyn Bartram, Scott Bateman, Sheelagh Carpendale, Anthony Tang, and Robert Woodbury. 2014. Personal visualization and personal visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3 (2014), 420–433.

[121] Theodore J Huppert, Solomon G Diamond, Maria A Franceschini, and David A Boas. 2009. HomER: a review of time-series analysis

methods for near-infrared spectroscopy of the brain. *Applied optics* 48, 10 (2009), D280–D298.

[122] Theodore J Huppert, Rick D Hoge, Solomon G Diamond, Maria Angela Franceschini, and David A Boas. 2006. A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29, 2 (2006), 368–382.

[123] Vanessa Ibáñez, Josep Silva, and Omar Cauli. 2018. A survey on sleep questionnaires and diaries. *Sleep medicine* 42 (2018), 90–96.

[124] Marcello Ienca and Roberto Andorno. 2017. Towards new human rights in the age of neuroscience and neurotechnology. *Life sciences, society and policy* 13, 1 (2017), 1–27.

[125] Marcello Ienca, Joseph J Fins, Ralf J Jox, Fabrice Jotterand, Silja Voeneky, Roberto Andorno, Tonio Ball, Claude Castelluccia, Ricardo Chavarriaga, Hervé Chneiweiss, et al. 2021. Towards a Governance Framework for Brain Data. *arXiv preprint arXiv:2109.11960* (2021).

[126] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 93–102. https://doi.org/10.1145/1357054.1357070

[127] Herbert Jasper. 1958. Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr Clin Neurophysiol* 10 (1958), 370–375.

[128] Ryan C Johnson, Erin M Eatough, Leslie B Hammer, Donald Truxillo, et al. 2019. Home is where the mind is: Family interference with

work and safety performance in two high risk industries. *Journal of vocational behavior* 110 (2019), 117–130.

[129] PGAM Jorna. 1993. Heart rate and workload variations in actual and simulated flight. *Ergonomics* 36, 9 (1993), 1043–1054.

[130] Marcel Adam Just, Patricia A Carpenter, Timothy A Keller, William F Eddy, and Keith R Thulborn. 1996. Brain activation modulated by sentence comprehension. *Science* 274, 5284 (1996), 114–116.

[131] Marcel Adam Just, Patricia A Carpenter, and Akira Miyake. 2003. Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science* 4, 1-2 (2003), 56–88.

[132] Daniel Kahneman. 1973. *Attention and effort*. Vol. 1063. Citeseer.

[133] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, and Mo Nours Arab. 2017. Human-computer interaction: Overview on state of the art. *International journal on smart sensing and intelligent systems* 1, 1 (2017).

[134] Kota Katanoda, Kohki Yoshikawa, and Morihiro Sugishita. 2001. A functional MRI study on the neural substrates for writing. *Human brain mapping* 13, 1 (2001), 34–42.

[135] Elisabeth T Kersten-van Dijk, Joyce HDM Westerink, Femke Beute, and Wijnand A IJsselsteijn. 2017. Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human–Computer Interaction* 32, 5-6 (2017), 268–296.

[136] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to evaluate technologies for health behavior change in HCI research. In

Proceedings of the SIGCHI conference on human factors in computing systems. 3063–3072.

[137] Li-Wei Ko, Oleksii Komarov, W David Hairston, Tzyy-Ping Jung, and Chin-Teng Lin. 2017. Sustained attention in real classroom settings: An EEG study. *Frontiers in human neuroscience* 11 (2017), 388.

[138] Arthur F Kramer. 2020. Physiological metrics of mental workload: A review of recent progress. *Multiple-task performance* (2020), 279–328.

[139] Karola V Kreitmair. 2019. Dimensions of ethical direct-to-consumer neurotechnologies. *AJOB neuroscience* 10, 4 (2019), 152–166.

[140] Karola V Kreitmair, Mildred K Cho, and David C Magnus. 2017. Consent and engagement, security, and authentic living using wearable and mobile health technology. *Nature biotechnology* 35, 7 (2017), 617–620.

[141] Simon Ladouce, David I Donaldson, Paul A Dudchenko, and Magdalena Ietswaart. 2017. Understanding minds in real-world environments: toward a mobile cognition approach. *Frontiers in human neuroscience* 10 (2017), 694.

[142] Richard S Lazarus. 2006. *Stress and emotion: A new synthesis*. Springer publishing company.

[143] Mikhail A Lebedev and Miguel AL Nicolelis. 2006. Brain–machine interfaces: past, present and future. *TRENDS in Neurosciences* 29, 9 (2006), 536–546.

[144] Kathryn A Lee, Gregory Hicks, and German Nino-Murcia. 1991. Va-

lidity and reliability of a scale to assess fatigue. *Psychiatry research* 36, 3 (1991), 291–298.

[145] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 557–566.

[146] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing*. 405–414.

[147] Chin-Teng Lin, Shu-Fang Tsai, and Li-Wei Ko. 2013. EEG-based learning system for online motion sickness level estimation in a dynamic vehicle environment. *IEEE transactions on neural networks and learning systems* 24, 10 (2013), 1689–1700.

[148] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*. Springer, 261–278.

[149] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How WEIRD is CHI?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[150] Vincent Lubrano, Franck-Emmanuel Roux, and Jean-François Démonet. 2004. Writing-specific sites in frontal areas: a cortical stimulation study. *Journal of neurosurgery* 101, 5 (2004), 787–798.

[151] Kristiyan Lukanov, Horia A Maior, and Max L Wilson. 2016. Using fNIRS in usability testing: understanding the effect of web form lay-

out on mental workload. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4011–4016.

[152] Horia A Maior, Matthew Pike, Sarah Sharples, and Max L Wilson. 2015. Examining the reliability of using fNIRS in realistic hci settings for spatial and verbal tasks. In *Proceedings of CHI*, Vol. 15. 3807–3816.

[153] Horia A Maior, Max L Wilson, and Sarah Sharples. 2018. Workload alerts-using physiological measures of mental workload to provide feedback during tasks. *ACM Transactions on Computer-Human Interaction* 25, 2 (2018).

[154] Gloria Mark, Victor M Gonzalez, and Justin Harris. 2005. No task left behind?: examining the nature of fragmented work. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 321–330.

[155] Gloria Mark, Daniela Gudith, and Ulrich Klocke. 2008. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 107–110.

[156] Gloria Mark, Shamsi Iqbal, and Mary Czerwinski. 2017. How blocking distractions affects workplace focus and productivity. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 928–934.

[157] David Marshall, Damien Coyle, Shane Wilson, and Michael Callaghan. 2013. Games, gameplay, and BCI: the state of the art. *IEEE Transactions on Computational Intelligence and AI in Games* 5, 2 (2013), 82–99.

[158] Pawel J Matusz, Suzanne Dikker, Alexander G Huth, and Catherine Perrodin. 2019. Are we ready for real-world neuroscience?

[159] Ted Megaw. 2005. The definition and measurement of mental workload. *Evaluation of human work, Eds. Esmond N. Corlett, and John R. Wilson* (2005), 525–551.

[160] Serena Midha, Horia A Maior, Max L Wilson, and Sarah Sharples. 2021. Measuring Mental Workload Variations in Office Work Tasks using fNIRS. *International Journal of Human-Computer Studies* 147 (2021), 102580.

[161] Earl K Miller and Jonathan D Cohen. 2001. An integrative theory of prefrontal cortex function. *Annual review of neuroscience* 24, 1 (2001), 167–202.

[162] Neville Moray. 2013. *Mental workload: Its theory and measurement.* Vol. 8. Springer Science & Business Media.

[163] Karim Moustafa, Saturnino Luz, and Luca Longo. 2017. Assessment of mental workload: a comparison of machine learning methods and subjective assessment techniques. In *International symposium on human mental workload: Models and applications*. Springer, 30–50.

[164] Frederick A Muckler and Sally A Seven. 1992. Selecting performance measures:" Objective" versus" subjective" measurement. *Human factors* 34, 4 (1992), 441–455.

[165] Takao Nakagawa, Yasutaka Kamei, Hidetake Uwano, Akito Monden, Kenichi Matsumoto, and Daniel M German. 2014. Quantifying programmers' mental workload during program comprehension based on cerebral blood flow measurement: a controlled experiment. In *Com-*

panion Proceedings of the 36th International Conference on Software Engineering. ACM, 448–451.

[166] Noman Naseer, Melissa Jiyoun Hong, and Keum-Shik Hong. 2014. Online binary decision decoding using functional near-infrared spectroscopy for the development of brain–computer interface. *Experimental brain research* 232, 2 (2014), 555–564.

[167] Adam R Nicholls, Remco CJ Polman, and Andrew R Levy. 2012. A path analysis of stress appraisals, emotions, coping, and performance satisfaction among athletes. *Psychology of sport and exercise* 13, 3 (2012), 263–270.

[168] Anton Nijholt, Desney Tan, Brendan Allison, Jose del R. Milan, and Bernhard Graimann. 2008. Brain-computer interfaces for HCI and games. In *CHI'08 extended abstracts on Human factors in computing systems*. 3925–3928.

[169] Ilse MT Nijs, Ingmar HA Franken, and Peter Muris. 2007. The modified Trait and State Food-Cravings Questionnaires: development and validation of a general index of food craving. *Appetite* 49, 1 (2007), 38–46.

[170] Francisco Nunes, Nervo Verdezoto, Geraldine Fitzpatrick, Morten Kyng, Erik Grönvall, and Cristiano Storni. 2015. Self-care technologies in HCI: Trends, tensions, and opportunities. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 6 (2015), 1–45.

[171] Brid O'Conaill and David Frohlich. 1995. Timespace in the workplace: Dealing with interruptions. In *Conference companion on Human factors in computing systems*. ACM, 262–263.

[172] Robert D O'Donnell and F Thomas Eggemeier. 1986. Workload assessment methodology. (1986).

[173] Kathryn M Orzech, Michael A Grandner, Brandy M Roane, and Mary A Carskadon. 2016. Digital media use in the 2 h before bedtime is associated with sleep variables in university students. *Computers in human behavior* 55 (2016), 43–50.

[174] Oskar Palinko, Andrew L Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications*. 141–144.

[175] Natacha Paquette, Maryse Lassonde, Phetsamone Vannasing, Julie Tremblay, Berta González-Frankenberger, Olivia Florea, Renée Béland, Franco Lepore, and Anne Gallagher. 2015. Developmental patterns of expressive language hemispheric lateralization in children, adolescents and adults using functional near-infrared spectroscopy. *Neuropsychologia* 68 (2015), 117–125.

[176] Raja Parasuraman. 2003. Neuroergonomics: Research and practice. *Theoretical issues in ergonomics science* 4, 1-2 (2003), 5–20.

[177] Raja Parasuraman. 2011. Neuroergonomics: Brain, cognition, and performance at work. *Current directions in psychological science* 20, 3 (2011), 181–186.

[178] Raja Parasuraman and Matthew Rizzo. 2008. *Neuroergonomics: The brain at work*. Oxford University Press.

[179] Raja Parasuraman and Glenn F Wilson. 2008. Putting the brain to work: Neuroergonomics past, present, and future. *Human factors* 50, 3 (2008), 468–474.

[180] Mitesh S Patel, David A Asch, and Kevin G Volpp. 2015. Wearable devices as facilitators, not drivers, of health behavior change. *Jama* 313, 5 (2015), 459–460.

[181] Jessica L Paterson, Jillian Dorrian, Jan Pincombe, Carol Grech, and Drew Dawson. 2010. Mood change and perception of workload in Australian midwives. *Industrial health* 48, 4 (2010), 381–389.

[182] Evan M Peck, Beste F Yuksel, Alvitta Ottley, Robert JK Jacob, and Remco Chang. 2013. Using fNIRS Brain Sensing to Evaluate Information Visualization Interfaces. In *Proc. SIGCHI*. ACM.

[183] F Pichiorri, F De Vico Fallani, F Cincotti, F Babiloni, M Molinari, SC Kleih, C Neuper, A Kübler, and D Mattia. 2011. Sensorimotor rhythm-based brain–computer interface training: the impact on motor cortical responsiveness. *Journal of neural engineering* 8, 2 (2011), 025020.

[184] Laura Pickup, John R Wilson, Beverley J Norris, Lucy Mitchell, and Ged Morrisroe. 2005. The Integrated Workload Scale (IWS): a new self-report tool to assess railway signaller workload. *Applied Ergonomics* 36, 6 (2005), 681–693.

[185] Matthew F Pike, Horia A Maior, Martin Porcheron, Sarah C Sharples, and Max L Wilson. 2014. Measuring the effect of think aloud protocols on workload using fNIRS. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 3807–3816.

[186] Paola Pinti, Clarisse Aichelburg, Sam Gilbert, Antonia Hamilton, Joy Hirsch, Paul Burgess, and Ilias Tachtsidis. 2018. A review on the use of wearable functional near-infrared spectroscopy in naturalistic

environments. *Japanese Psychological Research* 60, 4 (2018), 347–373.

[187] Paola Pinti, Clarisse Aichelburg, Sam Gilbert, Antonia Hamilton, Joy Hirsch, Paul Burgess, and Ilias Tachtsidis. 2018. A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Japanese Psychological Research* 60, 4 (2018), 347–373.

[188] Paola Pinti, Felix Scholkmann, Antonia Hamilton, Paul Burgess, and Ilias Tachtsidis. 2018. Current status and issues regarding pre-processing of fNIRS neuroimaging data: An investigation of diverse signal filtering methods within a General Linear Model framework. *Frontiers in human neuroscience* 12 (2018), 505.

[189] Paola Pinti, Ilias Tachtsidis, Antonia Hamilton, Joy Hirsch, Clarisse Aichelburg, Sam Gilbert, and Paul W Burgess. 2018. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences* (2018).

[190] Marc N Potenza and Carlos M Grilo. 2014. How relevant is food craving to obesity and its treatment? *Frontiers in psychiatry* 5 (2014), 164.

[191] Zachary Pousman, John Stasko, and Michael Mateas. 2007. Casual information visualization: Depictions of data in everyday life. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1145–1152.

[192] Jonathan D Power, Anish Mitra, Timothy O Laumann, Abraham Z Snyder, Bradley L Schlaggar, and Steven E Petersen. 2014. Methods

to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84 (2014), 320–341.

[193] Cathy J Price. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62, 2 (2012), 816–847.

[194] Amon Rapp and Maurizio Tirassa. 2017. Know thyself: a theory of the self for personal informatics. *Human–Computer Interaction* 32, 5-6 (2017), 335–380.

[195] Gary B Reid and Thomas E Nygren. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. *Advances in psychology* 52 (1988), 185–218.

[196] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1163–1172.

[197] Alan H Roscoe. 1992. Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological psychology* 34, 2-3 (1992), 259–287.

[198] Marjan Saadati, Jill Nelson, and Hasan Ayaz. 2019. Convolutional neural network for hybrid fNIRS-EEG mental workload classification. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 221–232.

[199] Angelo Sassaroli, Feng Zheng, Leanne M Hirshfield, Audrey Girouard, Erin Treacy Solovey, Robert JK Jacob, and Sergio Fantini. 2008. Discrimination of mental workload levels in human sub-

jects with functional near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences* 1, 02 (2008), 227–237.

[200] Walter Schneider and Richard M Shiffrin. 1977. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological review* 84, 1 (1977), 1.

[201] Felix Scholkmann, Stefan Kleiser, Andreas Jaakko Metz, Raphael Zimmermann, Juan Mata Pavia, Ursula Wolf, and Martin Wolf. 2014. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85 (2014), 6–27.

[202] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro. 2012. *STOP, THAT and one hundred other sleep scales*. Springer Science & Business Media.

[203] Sarah Sharples. 2018. Workload II: A Future Paradigm for Analysis and Measurement. In *Congress of the International Ergonomics Association*. Springer, 489–498.

[204] Sarah Sharples and Ted Megaw. 2015. Definition and Measurement of Human Workload. In *Evaluation of human work*, John R Wilson and Sarah Sharples (Eds.). CRC Press.

[205] Ben Shneiderman and Catherine Plaisant. 2005. Designing the User Interface: Strategies for Effective Human Computer Interaction. *Addison-Wesley, Reading, Mass* 4th edn (2005).

[206] Jonathan A Smith. 2004. Reflecting on the development of interpretative phenomenological analysis and its contribution to qualitative research in psychology. *Qualitative research in psychology* 1, 1 (2004), 39–54.

[207] Jonathan A Smith and Pnina Shinebourne. 2012. *Interpretative phenomenological analysis.* American Psychological Association.

[208] Erin Solovey, Paul Schermerhorn, Matthias Scheutz, Angelo Sassaroli, Sergio Fantini, and Robert Jacob. 2012. Brainput: enhancing interactive systems with streaming fNIRs brain input. In *CHI*. ACM, 2193–2202.

[209] Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert JK Jacob. 2009. Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In *Proc. UIST*. ACM, 157–166.

[210] Katarzyna Stawarz, Anna L. Cox, and Ann Blandford. 2015. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. In *Proc. CHI'15* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2653–2662. https://doi.org/10.1145/2702123.2702230

[211] Desney Tan and Anton Nijholt. 2010. Brain-computer interfaces and human-computer interaction. In *Brain-Computer Interfaces*. Springer, 3–19.

[212] Kazuo Tanaka, Kazuyuki Matsunaga, and Hua O Wang. 2005. Electroencephalogram-based control of an electric wheelchair. *IEEE transactions on robotics* 21, 4 (2005), 762–766.

[213] Jaime Teevan, Shamsi T Iqbal, Carrie J Cai, Jeffrey P Bigham, Michael S Bernstein, and Elizabeth M Gerber. 2016. Productivity decomposed: Getting big things done with little microtasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 3500–3507.

[214] Peter C Terry, Andrew M Lane, and Gerard J Fogarty. 2003. Construct validity of the Profile of Mood States—Adolescents for use with adults. *Psychology of sport and exercise* 4, 2 (2003), 125–139.

[215] D Tomasi, L Chang, EC Caparelli, and T Ernst. 2007. Different activation patterns for working memory load and visual attention load. *Brain research* 1132 (2007), 158–165.

[216] UNESCO International Bioethics Committee. 2021. Draft report of the IBC on ethical issues of neurotechnology. *UNESCO* SHS/IBC-28/2021/3 (Aug 2021), 52. https://unesdoc.unesco.org/ark:/48223/pf0000378724

[217] Bram van de Laar, Hayrettin Gürkök, Danny Plass-Oude Bos, Mannes Poel, and Anton Nijholt. 2013. Experiencing BCI control in a popular computer game. *IEEE Transactions on Computational Intelligence and AI in Games* 5, 2 (2013), 176–184.

[218] Giovanni Vecchiato, Gianluca Di Flumeri, Anton Giulio Maglione, Patrizia Cherubino, Wanzeng Kong, Arianna Trettel, and Fabio Babiloni. 2014. An electroencephalographic Peak Density Function to detect memorization during the observation of TV commercials. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 6969–6972.

[219] JA Veltman and AWK Gaillard. 1993. Indices of mental workload in a complex task environment. *Neuropsychobiology* 28, 1-2 (1993), 72–75.

[220] Grégoire Vergotte, Stéphane Perrey, Muthuraman Muthuraman, Stefan Janaqi, and Kjerstin Torre. 2018. Concurrent changes of brain functional connectivity and motor variability when adapting to task constraints. *Frontiers in physiology* 9 (2018), 909.

[221] Jacques J Vidal. 1973. Toward direct brain-computer communication. (1973).

[222] Jose Vina, F Sanchis-Gomar, V Martinez-Bello, and MC Gomez-Cabrera. 2012. Exercise acts as a drug; the pharmacological benefits of exercise. *British journal of pharmacology* 167, 1 (2012), 1–12.

[223] Hongtao Wang, Andrei Dragomir, Nida Itrat Abbasi, Junhua Li, Nitish V Thakor, and Anastasios Bezerianos. 2018. A novel real-time driving fatigue detection system based on wireless dry EEG. *Cognitive neurodynamics* 12, 4 (2018), 365–376.

[224] Shouyi Wang, Jacek Gwizdka, and W Art Chaovalitwongse. 2015. Using wireless EEG signals to assess memory workload in the n-back task. *IEEE Transactions on Human-Machine Systems* 46, 3 (2015), 424–435.

[225] Shouyi Wang, Yiqi Zhang, Changxu Wu, Felix Darvas, and Wanpracha Art Chaovalitwongse. 2014. Online prediction of driver distraction based on brain activity patterns. *IEEE Transactions on Intelligent Transportation Systems* 16, 1 (2014), 136–150.

[226] Chun-Shu Wei, Yuan-Pin Lin, Yu-Te Wang, Chin-Teng Lin, and Tzyy-Ping Jung. 2018. A subject-transfer framework for obviating inter-and intra-subject variability in EEG-based drowsiness detection. *NeuroImage* 174 (2018), 407–419.

[227] Alan Traviss Welford. 1968. Fundamentals of skill. (1968).

[228] Anna Wexler and Peter B Reiner. 2019. Oversight of direct-to-consumer neurotechnologies. *Science* 363, 6424 (2019), 234–235.

[229] Harold Thomas Anthony Whiting. 1969. *Acquiring ball skill: A psychological interpretation.* Lea & Febiger.

[230] Christopher D Wickens. 2002. Multiple resources and performance prediction. *Theoretical issues in ergonomics science* 3, 2 (2002), 159–177.

[231] Christopher D Wickens. 2008. Multiple resources and mental workload. *The Journal of the Human Factors and Ergonomics Society* 50, 3 (2008), 449–455.

[232] Christopher D Wickens, Justin G Hollands, Simon Banbury, and Raja Parasuraman. 2015. *Engineering psychology & human performance*. Psychology Press.

[233] Hollands Justin G Wickens, Christopher D. 2000. Attention, time-sharing, and workload, Ch. 11. *Engineering psychology and human performance* (2000).

[234] Glenn F Wilson. 2002. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology* 12, 1 (2002), 3–18.

[235] James R Wilson and Jane A Rajan. 1995. Human-machine interfaces for systems control. *Evaluation of human work: a practical ergonomics methodology* (1995), 357–405.

[236] Max L Wilson, Serena Midha, Horia A Maior, Anna L Cox, Lewis L Chuang, and Lachlan D Urquhart. 2022. SIG: Moving from Brain-Computer Interfaces to Personal Cognitive Informatics. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.

[237] Max L Wilson, Natalia Sharon, Horia A Maior, Serena Midha, Michael P Craven, and Sarah Sharples. 2018. Mental workload as

personal data: designing a cognitive activity tracker. In *Proceedings of the 3rd Symposium on Computing and Mental Health.*

[238] Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. 2002. Brain–computer interfaces for communication and control. *Clinical neurophysiology* 113, 6 (2002), 767–791.

[239] Nechama Yehuda. 2011. Music and stress. *Journal of Adult Development* 18, 2 (2011), 85–94.

[240] Mark S Young, Karel A Brookhuis, Christopher D Wickens, and Peter A Hancock. 2015. State of science: mental workload in ergonomics. *Ergonomics* 58, 1 (2015), 1–17.

[241] Mark S Young and Neville A Stanton. 2002. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human factors* 44, 3 (2002), 365–375.

[242] Beste F Yuksel, Kurt B Oleson, Lane Harrison, Evan M Peck, Daniel Afergan, Remco Chang, and Robert JK Jacob. 2016. Learn piano with BACh: An adaptive learning interface that adjusts task difficulty based on brain state. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 5372–5384.

[243] Thorsten O Zander, Jonas Brönstrup, Romy Lorenz, and Laurens R Krol. 2014. Towards BCI-based implicit control in human–computer interaction. In *Advances in Physiological Computing.* Springer, 67–90.

[244] Thorsten O Zander and Christian Kothe. 2011. Towards passive brain–computer interfaces: applying brain–computer interface tech-

nology to human–machine systems in general. *Journal of neural engineering* 8, 2 (2011), 025005.

[245] Thorsten O Zander, Christian Kothe, Sabine Jatzev, and Matti Gaertner. 2010. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-computer interfaces*. Springer, 181–199.

[246] Yibo Zhu, Carolina Rodriguez-Paras, Joohyun Rhee, and Ranjana K Mehta. 2020. Methodological approaches and recommendations for functional near-infrared spectroscopy applications in HF/E research. *Human factors* 62, 4 (2020), 613–642.

Appendices

Appendix A

Study 2 and 3 Documents

A.1 Information Sheet

PROJECT INFORMATION

University of Nottingham
UK | CHINA | MALAYSIA

Date: 12/06/2020

Project: Low Density Data Collection

School of Computer Science Ethics Reference: [CS-2019-R13]

Funded by: Horizon CDT, Mixed Reality Laboratory

We would like to invite you to take part in our research study. Before you decide whether to participate, we would like you to understand why the research is being done and what it would involve for you. One of our team will go through the information sheet with you and answer any questions you have. Talk to others about the study if you wish. Ask us if there is anything that is not clear.

What is the purpose of the study?

This research project aims to investigate individual variations of mental workload levels in daily life, how mental workload is considered, and what consequences tracking mental workload might have, what impact this may have, and the individual experiences of this.

Why have I been invited?

You are being invited to take part because you are considered to represent a profile which might have use in the future for a device that could track brain activity, like a Fitbit. We are inviting 20 participants like you to take part.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. This would not affect your legal rights.

What will happen to me if I take part?

You will be required to be a participant in the study for 5 full days. Throughout this time, you will be asked to continue with your activities as normal with the addition of completing certain measures and allowing further measures to be taken from you. Before beginning the study, you will be asked to prepare for the data collection by reading a pre-study preparation sheet. At a specified time the week following the data collection, you will be asked to take part in a post-study interview. Each measure is outlined below; for the measures that you need to install, you will be guided through how to do this:

Mobile application: We will ask you to install a file onto your phone that includes an application that is not commercially available. Every 1 hour outside of working hours, and every 30 minutes during working hours, you will be asked to answer 3 questions through a mobile application; a notification will alert you when your responses are due. Each question will require one button click and should only take about 10 seconds to complete. At the same time, a final question will ask you to complete a brief text summary of the activity/activities you have participated in since the last input; if there is information that you do not wish to share, please mention that you are withholding information, or provide the nature of the information you do not wish to share. You are encouraged to complete these questions as soon as you are alerted, but the time for the next notification will begin after your last completion. You should activate the app when you wake up each day and deactivate it before going to sleep; compliance scores will only be counted during waking hours, and the practicalities sheet will detail how to activate and deactivate the app.

Evening Questionnaires: Each evening you will be asked to complete a further 7 questionnaires. These questionnaires should be completed using Microsoft Word and should take approximately 15 minutes to complete. The questionnaires will regard your stress level, sleep, fatigue, mood, food craving, food intake, and alcohol consumption. You will be asked to complete them at approximately the same time each evening – this should be *after* your evening meal, and preferably close to bedtime. We will also ask you to complete a personality and personal information questionnaire, but this will only need to be completed once. If possible, the sleep questionnaire should be started on Tuesday (at any time convenient for you) and the last entry should be on Saturday, so there is data from Monday-Friday on your sleep; it might be convenient to complete the last sleep entry on Saturday when uninstalling the software.

Online activity: We will ask you to install an application called DeskTime onto your computer. This will track your online activity. The application will be able to track which websites and programmes you are using, including the website or document titles, what time you are using them, and how long you are using them for. The application will not be able to track the content of what is on your screen; this information will only be accessible to you. If you wish to stop your online activity from being tracked at any time or remove any data, you will be provided with information about how to do this.

Phone activity: We will also ask you to install an application called RescueTime on your personal mobile device. The application will track what applications you are using, and what time and for how long you are using them for. RescueTime will not track what you are doing on the applications (e.g. which websites you are using); it will simply track what applications are open. If you wish to stop this activity being tracked at any time or remove any data, you will be provided with information about how to do this.

Calendar: you will be asked to provide your schedule for the 5 days of data collection, in any form. You are not required to provide any sensitive information; if there is information that you do not wish to share you may leave the space blank in your schedule, or provide the nature of the information, e.g. 'sensitive meeting'.

Post study interview: The interview may take up to 2 hours and will be taken via video call. In order to analyse the interview, it will be recorded.

Expenses and payments

You will be offered £75 to participate in the study. With satisfactory participation in the study, the amount will be increased to £100. Participation will be considered satisfactory if all evening questionnaires are completed, and there is at least an 80% response rate to the mobile application questions.

What are the possible disadvantages and risks of taking part?

Your safety and confidentiality are of the utmost importance to us, and these will not be compromised through your participation in this study. However, a possible disadvantage of participation may include the disruption of your activity when required to complete the mobile questions; whilst we have tried to make this measure as least intrusive as we can, it will still require input that is separate to your current activity.

What are the possible benefits of taking part?

We cannot promise the study will help you but the information we get from this study may help progress research into gaining a much deeper understanding of mental workload in an everyday context and identifying which measures can be used to capture mental workload levels in daily life.

What happens when the research study stops?

You will be able to return to your routine without any further requirements from the experimenters. You will always be welcome to contact the research team if you have any further queries after the study has ended. After the end of the study, the research team will spend time analysing the data that you have helped provide in order to answer our research questions. All of this data will be stored securely, remaining available to the research team for up to 10 years before being deleted. Data in which you cannot be identified from, such as digitalised questionnaire responses, may be made available on the University research dataset archive.

What if there is a problem?

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions. The researchers contact details are given at the end of this information sheet. If you remain unhappy and wish to complain formally, you can do this by the University.

Will my taking part in the study be kept confidential?

Your participation in the study will be kept strictly confidential before, during, and after data collection. Your personal data will be protected in line with the General Data Protection Regulation (GDPR) and the Data Protection Act 2018; further information about this is provided on a separate sheet. Please be assured that your data will be anonymised and you will not be identifiable in any data, analyses, or publications, and any data in which you may be identifiable (e.g. audio recordings) will only be accessible to the experimenters. We will follow ethical and legal practice and all information about you will be handled in confidence.

All data collected from you will comply with the GDPR guidelines. This includes DeskTime and RescueTime; the experimenters will delete your accounts for these applications after data analysis which will remove your data from their databases.

Please note that the experimenters have a duty of care and are obligated to act accordingly if any data collected indicates endangerment to the health or life of yourself or others, or indicates a criminal act. If this is the case, protocol will be first for the researcher to discuss the data with a senior member of the research team. For this study, if questionnaire data indicates excessive alcohol consumption, the researcher would discuss the findings with you and provide guidance on where to seek help. If the questionnaire data or online activity data indicates a clinically relevant low mood, the researcher would discuss the data with you and provide guidance on where to seek help, unless this is deemed to be more harmful to you. In this case, the researcher may raise their concern about you directly with a mental health professional. Although the data collected is unlikely to identify criminal activity, if contacted by authorities about the data then we shall cooperate.

What will happen if I don't want to carry on with the study?

Your participation is voluntary and you are free to withdraw at any time, without giving any reason, and without your legal rights being affected. If you withdraw then the information collected so far may still be used in the project analysis unless specifically requested otherwise. Once data analysis has started and your data has been anonymised, you can no longer withdraw your participation.

What will happen to the results of the research study?

The results of the research will be used as the basis for a PhD thesis submitted in 2021. The results are also likely to be published in conference papers and/or journal articles, and be discussed in research activities, such as presentations.

Who is organising and funding the research?

The research is being organised by the University of Nottingham and is being funded by the Horizon CDT and Mixed Reality Laboratory based in the School of Computer Science.

Who has reviewed the study?

All proposed research in the University is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and approved by the School of Computer Science Research Ethics Committee in the University of Nottingham.

Further information and contact details

If you have any further queries, please do not hesitate to contact the research team:

Serena Midha: serena.midha@nottingham.ac.uk

Dr Max L. Wilson: max.wilson@nottingham.ac.uk

Dr Horia Maior: horia.maior@nottingham.ac.uk

Contact details of the ethics committee.

If you wish to file a complaint or exercise your rights you can contact the Ethics Committee at the following address: cs-ethicsadmin@cs.nott.ac.uk

A.2 Privacy Notice

PRIVACY NOTICE

The University of Nottingham is committed to protecting your personal data and informing you of your rights in relation to that data. The University will process your personal data in accordance with the General Data Protection Regulation (GDPR) and the Data Protection Act 2018 and this privacy notice is issued in accordance with GDPR Articles 13 and 14.

The University of Nottingham, University Park, Nottingham, NG7 2RD is registered as a Data Controller under the Data Protection Act 1998 (registration No. Z5654762, https://ico.org.uk/ESDWebPages/Entry/Z5654762).

The University has appointed a Data Protection Officer (DPO). The DPO's postal address is:

Data Protection Officer,
Legal Services
A5, Trent Building,
University of Nottingham,
University Park,
Nottingham
NG7 2RD

The DPO can be emailed at dpo@nottingham.ac.uk

Why we collect your personal data. We collect personal data under the terms of the University's Royal Charter in our capacity as a teaching and research body to advance education and learning. Specific purposes for data collection on this occasion are to monitor and understand mental workload levels in daily life.

The legal basis for processing your personal data under GDPR. Under the General Data Protection Regulation, the University must establish a legal basis for processing your personal data and communicate this to you. The legal basis for processing your personal data on this occasion is Article 6(1e) processing is necessary for the performance of a task carried out in the public interest.

Special category personal data. In addition to the legal basis for processing your personal data, the University must meet a further basis when processing any special category data, including: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. The basis for processing your sensitive personal data on this occasion is Article 9(2j) processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.

How long we keep your data. The University may store your data for up to 25 years and for a period of no less than 7 years after the research project finishes. The researchers who gathered or processed the data may also store the data indefinitely and reuse it in future research.

Who we share your data with. Your data may be shared with researchers from other collaborating institutions and organisations who are involved in the research. Extracts of your data may be disclosed in published works that are posted online for use by the scientific community. Your data may also be stored indefinitely by members of the researcher team and/or be stored on external data repositories (e.g., the UK Data Archive) and be further processed for archiving purposes in the public interest, or for historical, scientific or statistical purposes.

How we keep your data safe. We keep your data securely and put measures in place to safeguard it. These safeguards include the anonymisation of all data, storage of digital data on password protected devices and external hard drives, secure storage of paper data kept on the University of Nottingham premesis, and stored on university protected servers.

Your rights as a data subject. GDPR provides you, as a data subject, with a number of rights in relation to your personal data. Subject to some exemptions, you have the right to:

- withdraw your consent at any time where that is the legal basis of our processing, and in such circumstances you are not obliged to provide personal data for our research.
- object to automated decision-making, to contest the decision, and to obtain human intervention from the controller.
- access (i.e., receive a copy of) your personal data that we are processing together with information about the purposes of processing, the categories of personal data concerned, recipients/categories of recipient, retention periods, safeguards for any overseas transfers, and information about your rights.
- have inaccuracies in the personal data that we hold about you rectified and, depending on the purposes for which your data is processed, to have personal incomplete data completed
- be forgotten, i.e., to have your personal data erased where it is no longer needed, you withdraw consent and there is no other legal basis for processing your personal data, or you object to the processing and there is no overriding legitimate ground for that processing.
- in certain circumstances, request that the processing of your personal data be restricted, e.g., pending verification where you are contesting its accuracy or you have objected to the processing.
- obtain a copy of your personal data which you have provided to the University in a structured, commonly used electronic form (portability), and to object to certain processing activities such as processing based on the University's or someone else's legitimate interests, processing in

the public interest or for direct marketing purposes. In the case of objections based on the latter, the University is obliged to cease processing.

- complain to the Information Commissioner's Office about the way we process your personal data.

If you require advice on exercising any of the above rights, please contact the University's data protection team: data-protection@nottingham.ac.uk

A.3 Consent Form

CONSENT FORM

University of Nottingham
UK | CHINA | MALAYSIA

Date:

Project: Low Density Data Collection

School of Computer Science Ethics Reference: CS-2019-R13

Funded by: Horizon CDT, Mixed Reality Lab

Please tick the appropriate boxes	Yes	No
1. Taking part in the study		
a) I have read and understood the project information sheet dated 12/06/2020, or it has been read to me. I have been able to ask questions about the study and my questions have been answered satisfactorily.	☐	☐
b) I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	☐	☐
c) I understand that my data cannot be withdrawn from the analysis once it has been made anonymous.	☐	☐
d) I understand that taking part in the study requires me to provide data and that this will involve questionnaires, online and phone activity data, calendar information, and interview data.	☐	☐
2. Use of my data in the study		
a) I understand that data which can identify me will not be shared beyond the project team.	☐	☐
b) I agree that the data provided by me may be used for the following purposes:		
– Presentation and discussion of the project and its results in research activities (e.g., in supervision sessions, project meetings, conferences).	☐	☐
– Publications and reports describing the project and its results.	☐	☐
– Dissemination of the project and its results, including publication of data on web pages and databases.	☐	☐
c) I give permission for my words to be quoted for the purposes described above.	☐	☐

Please tick the appropriate boxes	Yes	No
3. Reuse of my data		
a) I give permission for the data that I provide to be reused for the sole purposes of future research and learning.	☐	☐
b) I understand and agree that this may involve depositing my data in a data repository, which may be accessed by other researchers	☐	☐

4. Security of my data

a) I understand that safeguards will be put in place to protect my identity and my data during the research, and if my data is kept for future use. ☐ ☐

b) I confirm that a written copy of these safeguards has been given to me in the University's privacy notice, and that they have been described to me and are acceptable to me. ☐ ☐

c) I understand that no computer system is completely secure and that there is a risk that a third party could obtain a copy of my data. ☐ ☐

d) I understand that under certain circumstances in which data indicates criminal activity or endangerment to health or life, my data may be shared to relevant parties outside of the research team. ☐ ☐

5. Copyright

a) I give permission for data gathered during this project to be used, copied, excerpted, annotated, displayed and distributed for the purposes to which I have consented. ☐ ☐

6. Signatures (sign as appropriate)

______________________	______________________	__________
Name of participant (IN CAPITALS)	Signature	Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

______________________	______________________	__________
Name of researcher (IN CAPITALS)	Signature	Date

7. Researcher's contact details

Name: Serena Midha

Dr Max L. Wilson

Dr Horia Maior

Email: serena.midha@nottingham.ac.uk

max.wilson@nottingham.ac.uk

horia.maior@nottingham.ac.uk

A.4 Pre-Study Document

PRE-STUDY INFORMATION SESSION

To make sure we get the best out of your participation in the study, this sheet will explain it in more depth from what you might already know, not only to make sure you know what's coming up throughout the week, but also to start to get you to think about the study with a particular mindset and start to think about certain things a bit more deeply.

Study aims:

What we are researching is the concept of a Fitbit for the brain. The idea behind this is that we can track our physical states very effectively in our lives now, such as the amount of steps that we have walked or our heart rate. We can use that information to make improvements to our physical health, for example by aiming to walk a certain number of steps each day. Perhaps if we could track our cognitive activity in a similar way, we could have the opportunity to make further improvements to our lives.

The cognitive activity that we are interested in tracking is mental workload. There is actually no agreed upon definition of mental workload, meaning different people or researchers can interpret the meaning of it differently. Generally, though, it is agreed that the aspects mental workload is made up of are the demands of the task, the effort expended in responding to the task, and how well you are able to perform on the task. Mental workload is different to stress, as stress is the emotional strain an individual experiences from adverse or demanding circumstances; however, that is not to say that an individual can't experience both high mental workload and high stress when responding to a task, but it is important to differentiate between the two. We do also know that if your mental workload is too high at a certain time, your performance on the task will drop. This is perhaps why mental workload has so far only been investigated in the workplace.

But with variable levels of mental workload outside of the workplace, the work-life balance becoming ever more blurred, the impact that mental workload is thought to have on wellbeing, and a rising increase in the presence of technology, mental workload is also a meaningful concept outside of the workplace. So, we want to approach mental workload from a more holistic, life perspective, instead of just seeing it as something that is important in the workplace.

Because of this, we have identified three areas to investigate in this study:

1. Variations of mental workload in daily life and which measures can support this data.

2. How mental workload is conceptualised. – we want to develop a deeper understanding of your experiences and how you think about mental workload in order to inform design choices for the presentation of brain data.

3. How mental workload may be contextualised. – whilst mental workload is thought to be related to wellbeing, we're really interested in what specific things different levels of mental workload may affect.

What you'll have to do:

Your participation in the study will help us answer these questions. A more detailed description of what your participation will involve is found on the information sheet. But as an overview: from Monday-Friday we will collect a number of measures from you. These include online and phone activity data, your calendar, mental workload ratings with diary data, and questionnaire responses. We have designed the study to be as least intrusive to your life as possible, but effort will be required from you for certain measures. The following week, there will be a post-study interview about your experiences and views of mental workload.

What we want from you:

Aside from asking for your commitment to the study, we are creating the opportunity to have a very in depth look at mental workload. What we really want you to focus on in regard to your approach to the study is the conceptual experience of mental workload - in the post study interview, we will really probe you about your thoughts, experiences and views of mental workload. So to get the best out of the week, outlined below are the types of questions that we will be asking in the interview to start you thinking about what you are actually experiencing as the week goes on – this doesn't mean that you should feel the need to spend too much time thinking about these things, but they are just here to consider in order to shape your mindset. The interview will be in 4 parts:

Personal experiences: this will ask about your specific experiences of different mental workload levels throughout the week. You'll be probed about items such as how it felt to be at a high, medium, and low mental workload level, whether these feelings were the same regardless of the task, and how long you can maintain these levels for. We'll also be interested in examples of your experiences of dealing with mental workload; for example, someone previously explained that they had been experiencing high mental workload for such a long time, and all they could see in front of them was more high mental workload so they went to the toilet just because they thought they needed the relief of breaking up the mental workload level. Before the pandemic, another person explained that the process of leaving the house in the morning is the highest mental workload that they experience because they have so many things to remember; they even find themselves asking their daughter to be quiet because they would feel too overloaded with any more things on their mind. For a final example, someone else explained that they were going through a really tough week at work with a prolonged amount of high mental workload, and because of this they couldn't bring themselves to get on with other aspects of life (doing artwork) when they had spare time that week. So these types of personal experiences with mental workload are something we're really interested in and something that we'd like you to try and tune into during the week.

Conceptualisation: if we have an understanding about how you think about your mental workload, we could further understand mental workload from a life perspective and work on a way to present mental workload data back to you and other users effectively. So we'll be asking, for example, how you would define mental workload, and what kinds of metaphors and colours you would associate with different mental workload levels.

Contextualisation: as well as tracking mental workload in the real world, we're also interested in *why* people would want to track their mental workload – meaning, what are the consequences of mental workload? Perhaps you will notice what your mental workload levels appear to be affecting in, for example, your behaviour or cognition, and also what you would want to get from tracking your mental workload data.

Ethical considerations: we'll also touch upon whether you have any concerns about the concept of a device that could track your cognitive activity.

A.5 Practicalities Document

PRACTICALITIES

This sheet is designed to provide practical information about the measures being taken and situations you might encounter. Please do not hesitate to contact members of the research team if you have any queries that are not answered here:

Serena Midha: serena.midha@nottingham.ac.uk
Dr Max L. Wilson: max.wilson@nottingham.ac.uk
Dr Horia Maior: horia.maior@nottingham.ac.uk

Mobile Application:

There are four parts to the mobile app:

1. 'What is your current mental workload level?'
2. 'What has your overall mental workload level been since your last rating?'
3. 'What would you rate your overall performance since the last rating?'
4. 'Please use the space below to report a summary of the tasks performed since the last rating.'

The first two questions refer to your mental workload levels. The answer options range from Under-utilised to Excessive. The key to answering these questions accurately is outlined in the table below. Please note that this scale is a widely used scale and, like all available mental workload scales, has been designed with working tasks in mind; therefore, you may be required to make logical decisions when rating mental workload levels for activities outside of working tasks. In mental workload terms, spare capacity refers to the amount of mental resources that you have spare aside from the mental resources being used on the task. E.g. if you rate your mental workload level as Under-utilised, you should feel that you have very many spare mental resources that are not being allocated to the activity/activities that you are doing/have done. Please note in the description box, you are referred to as the controller. For question 1, you should rate your mental workload level based on how you feel at the current time of the rating. For question 2, you should rate your mental workload level based on your average experience since you last rated.

The third question has a scale ranging from very poor to very good. For this question, your rating should reflect whether you feel as though you have achieved what you were meant to. E.g. you could rate well in performance if you have been watching TV to relax and have been relaxing, but rate lower in performance if you haven't been able to relax.

The fourth question can be answered in as much detail as you deem appropriate; the more detail, the better the data will be for the researchers.

It may be most efficient to use one line for each activity that you enter. If you wish to withhold certain activities, please mention 'activity withheld'.

Mental Workload	Spare Capacity	Description
Under-utilised	Very much	Little or nothing to do. Rather boring.
Relaxed	Ample	More time than necessary to complete the tasks. Time passes slowly.
Comfortable	Some	The controller has enough work to keep him/her stimulated. All tasks are under control.
High	Very little	Certain non-essential tasks are postponed. Could not work at this level very long. Controller is working 'at the limit'. Time passes quickly.
Excessive	None	Some tasks are not completed. The controller is overloaded and does not feel in control.

When starting the app for the first time, you need to enter your working hours by selecting the 'working hours setup'. If your working hours are 9am to 5pm, that should be entered as 09 and 17. Once you have entered these details, they will be saved in the application so don't need to be changed unless your working hours change. Once entered, go back to the home screen and select 'start' to begin your ratings. When the app is activated each day, the first rating will only include the first question. When you finish your ratings, do not click 'stop rating' as this will deactivate the app – when you are taken to that screen, your rating has been logged and you can now leave the app until the next notification.

When you receive a notification, it is important to access the app through clicking on the notification. If you access the app directly instead of through the notification, the app will be deactivated and you will need to begin the ratings again. If this happens, don't worry and just begin again.

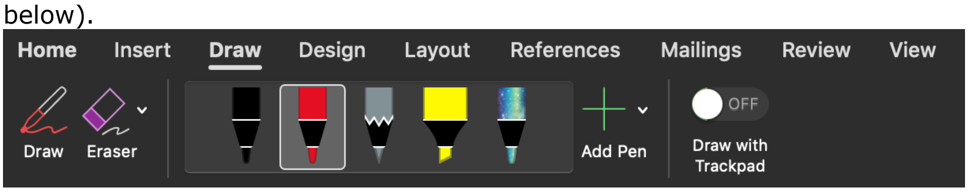
The app should be turned on from the time you wake up to the time you go to sleep. When you wish to turn the app off before sleep, click the 'stop rating' button. When you wake, please remember to re-start the app.

If you wish to change how you answered the questions on the mobile application, such as adding an activity or editing a rating, you can do so by getting in touch with the experimenter. Please send an email detailing what you would like to change, and the approximate time that you submitted your response.

Questionnaires

The questionnaires are provided in a Microsoft Word format. Please don't forget to save your responses. You may send completed questionnaires to the experimenter at any time during the study, whether this is each day, or all together at the end of the week.

When filling out the questionnaires, please indicate your selected answers in any way that you find suitable. A suggestion for this is to select the 'Draw' option, where you are free to mark and erase your answers freely (pictured below).

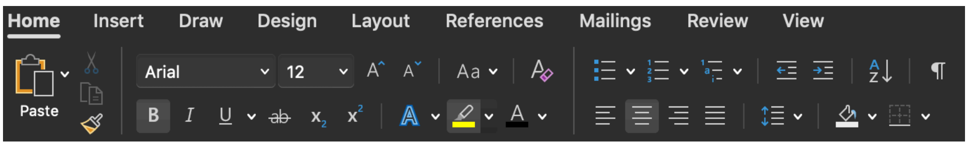


0 1 2 3 4

0 1 2 3 4

0 1 2 3 4

Other suggestions are to use the highlight option or bold text option, pictured below.



0 1 **2** 3 4	0 1 2 3 4
0 1 **2** 3 4	0 1 2 3 4
0 1 **2** 3 4	0 1 2 3 4

Deleting data

As mentioned in the information sheet, you will be able to delete data that you do not wish to be analysed. Below explains how this may be done:

DeskTime: as mentioned in the information sheet, the online activity tracker will not be able to track the content of what is on your screen. If you would like to stop the tracking of which websites and programmes you are using, including the website or document titles, what time you are using them, and how long you are using them for, you may temporarily disable the tracker. On a Mac, click on the DeskTime icon (a circle) at top righthand side of your screen and click on 'Private time'. Once you are happy for this data to be tracked again, please remember to click on 'Private time' once again to re-enable the tracker. For Windows, *right click* on the DeskTime icon (a circle clock filled with mostly green) and click on 'Private time'. Once you are happy for this data to be tracked again, please remember to click on 'Private time' once again to re-enable the tracker.

You can delete sections of time retrospectively by going onto the DeskTime interface. Navigate to the relevant date using the calendar icon, then select a blue horizontal section at the bottom of the 'Productivity Bar' graph. Adjust the time frame in which you wish to delete and then select delete.

RescueTime: If you would like to pause this activity from being tracked, you may disable the tracker. In the app, go to Settings and select 'pause tracking'. Please remember to uncheck this box once you are happy for this data to be tracked again.

You cannot remove data once it has been collected through the mobile app. If you would like to remove data retrospectively, you can ask the impartial experimenter to do this for you or remove it yourself through the desktop interface. To do this, download the RescueTime app onto your computer. Log in using the details you entered into the mobile app. From the desktop interface, navigate to the relevant day, click 'details' next to the phone icon, and hover the mouse over the activity or activities that you wish to remove. Then select the square icon with the pen in the middle which appears when you hover the mouse. Select delete and choose the relevant option.

A.6 Participant Bio Questionnaire

Participant Information

Participant ID:

Participant Name:

Date of Birth:

Gender:

Handedness:

Currently: Working from home / working from office / other

Job/level of study and course or research area:

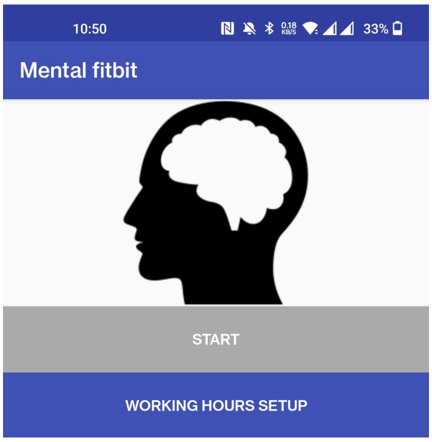
Summary of responsibilities:

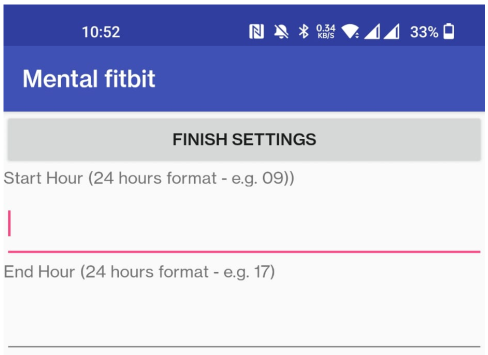
Usual working hours:

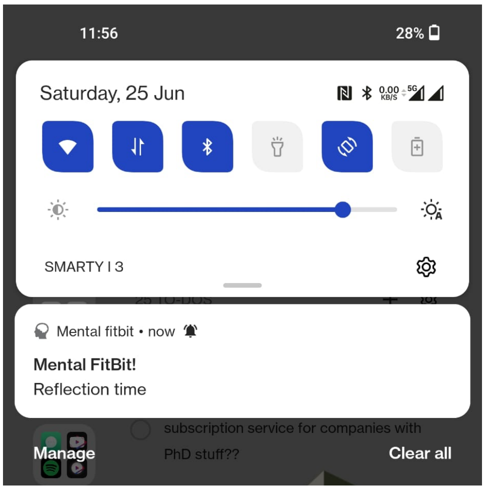
Appendix B

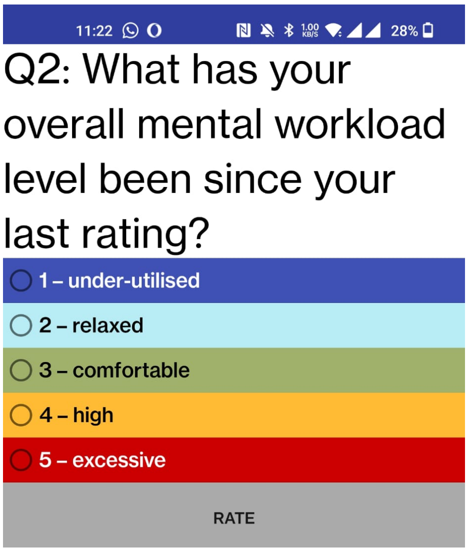
Study 2 Materials

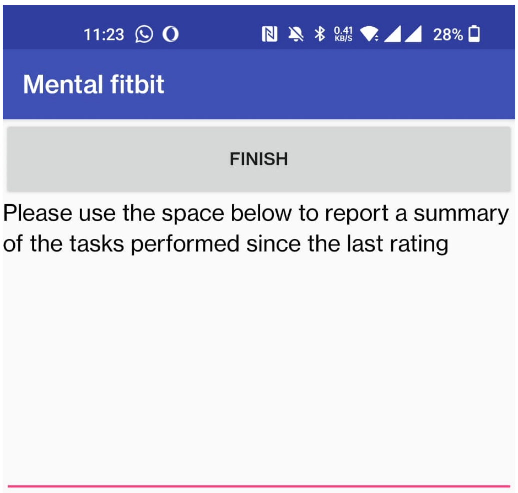
B.1 Ratings App











B.2 Perceived Stress Scale

PERCEIVED STRESS SCALE

Participant ID: **Date:** **Time:**

The questions in this scale ask you about your feelings and thoughts during the current day. In each case, you will be asked to indicate *how often* you felt or thought a certain way.

0=Never 1=Almost Never 2=Sometimes 3=Fairly Often 4=Very Often

1. Today, how often have you been upset because of something that happened unexpectedly?	0	1	**2**	3	4
2. Today, how often have you felt that you were unable to control the the important things in your life?	0	1	**2**	3	4
3. Today, how often have you felt nervous and “stressed”?	0	1	**2**	3	4
4. Today, how often have you felt confident about your ability to handle personal problems?	0	1	2	3	4
5. Today, how often have you felt that things were going your way?	0	1	2	3	4
6. Today, how often have you found that you could not cope with all the things that you had to do?	0	1	2	3	4
7. Today, how often have you been able to control irritations in your life?	0	1	2	3	4
8. Today, how often have you felt that you were on top of things?	0	1	2	3	4
9. Today, how often have you been angered because of things that were outside of your control?	0	1	2	3	4
10. Today, how often have you felt difficulties were piling up so high that you would not overcome them?	0	1	2	3	4

B.3 Revised ROMS

Abbreviated POMS (Revised Version)

Name: ______________________________ Date:

Below is a list of words that describe feelings people have. Please **CIRCLE THE NUMBER THAT BEST DESCRIBES HOW YOU FEEL RIGHT NOW**.

	Not At All	**A Little**	**Moderately**	**Quite a lot**	**Extremely**
Tense	0	1	2	3	4
Angry	0	1	2	3	4
Worn Out	0	1	2	3	4
Unhappy	0	1	2	3	4
Proud	0	1	2	3	4
Lively	0	1	2	3	4
Confused	0	1	2	3	4
Sad	0	1	2	3	4
Active	0	1	2	3	4
On-edge	0	1	2	3	4
Grouchy	0	1	2	3	4
Ashamed	0	1	2	3	4
Energetic	0	1	2	3	4
Hopeless	0	1	2	3	4
Uneasy	0	1	2	3	4
Restless	0	1	2	3	4
Unable to concentrate	0	1	2	3	4
Fatigued	0	1	2	3	4
Competent	0	1	2	3	4
Annoyed	0	1	2	3	4
Discouraged	0	1	2	3	4
Resentful	0	1	2	3	4
Nervous	0	1	2	3	4
Miserable	0	1	2	3	4

PLEASE CONTINUE WITH THE ITEMS ON THE NEXT PAGE

	Not At All	A Little	Moderately	Quite a lot	Extremely
Confident	0	1	2	3	4
Bitter	0	1	2	3	4
Exhausted	0	1	2	3	4
Anxious	0	1	2	3	4
Helpless	0	1	2	3	4
Weary	0	1	2	3	4
Satisfied	0	1	2	3	4
Bewildered	0	1	2	3	4
Furious	0	1	2	3	4
Full of Pep	0	1	2	3	4
Worthless	0	1	2	3	4
Forgetful	0	1	2	3	4
Vigorous	0	1	2	3	4
Uncertain about things	0	1	2	3	4
Bushed	0	1	2	3	4
Embarrassed	0	1	2	3	4

THANK YOU FOR YOUR COOPERATION

PLEASE BE SURE YOU HAVE ANSWERED EVERY ITEM

B.4 VAS-F

Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F)

Participant ID: **Date:** **Time:**

We are trying to find out about your level of energy. There are 18 items we would like you to respond to. This should take less than 1 minute of your time.

DIRECTIONS: You are asked to specify a number on each of the following lines to indicate how you are feeling RIGHT NOW.

For example, suppose you have not eaten since yesterday.
What number would you circle below?

not at all hungry 0 1 2 3 4 5 6 7 8 9 10 extremely hungry

You would probably specify a number closer to the “extremely hungry” end of the line. This is where I put it:

not at all hungry 0 1 2 3 4 5 6 7 8 9 10 extremely hungry

NOW PLEASE COMPLETE THE FOLLOWING ITEMS:

1. not at all **tired** 0 1 2 3 4 5 6 7 8 9 10 extremely **tired**
2. not at all **sleepy** 0 1 2 3 4 5 6 7 8 9 10 extremely **sleepy**
3. not at all **drowsy** 0 1 2 3 4 5 6 7 8 9 10 extremely **drowsy**
4. not at all **fatigued** 0 1 2 3 4 5 6 7 8 9 10 extremely **fatigued**
5. not at all **worn out** 0 1 2 3 4 5 6 7 8 9 10 extremely **worn out**
6. not at all **energetic** 0 1 2 3 4 5 6 7 8 9 10 extremely **energetic**
7. not at all **active** 0 1 2 3 4 5 6 7 8 9 10 extremely **active**

8. not at all **vigorous**	0 1 2 3 4 5 6 7 8 9 10	extremely **vigorous**
9. not at all **efficient**	0 1 2 3 4 5 6 7 8 9 10	extremely **efficient**
10. not at all **lively**	0 1 2 3 4 5 6 7 8 9 10	extremely **lively**
11. not at all **bushed**	0 1 2 3 4 5 6 7 8 9 10	extremely **bushed**
12. not at all **exhausted**	0 1 2 3 4 5 6 7 8 9 10	extremely **exhausted**
13. **keeping my eyes open** is no effort at all	0 1 2 3 4 5 6 7 8 9 10	**keeping my eyes open** is a tremendous chore
14. **moving my body** is no effort at all	0 1 2 3 4 5 6 7 8 9 10	**moving my body** is a tremendous chore
15. **concentrating** is no effort at all	0 1 2 3 4 5 6 7 8 9 10	**concentrating** is a tremendous chore
16. **carrying on a conversation** is no effort at all	0 1 2 3 4 5 6 7 8 9 10	**carrying on a conversation** is a tremendous chore
17. I have absolutely **no desire to close my eyes**	0 1 2 3 4 5 6 7 8 9 10	I have a tremendous **desire to close my eyes**
18. I have absolutely **no desire to lie down**	0 1 2 3 4 5 6 7 8 9 10	I have a tremendous **desire to lie down**

B.5 FCQ-S

Right now, at this very moment…		Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I have an intense desire to eat [one or more specific foods].	1	2	3	4	5
2	I’m craving [one or more specific foods].	1	2	3	4	5
3	I have an urge for [one or more specific foods].	1	2	3	4	5
4	Eating [one or more specific foods] would make things seem just perfect.	1	2	3	4	5
5	If I were to eat what I am craving, I am sure my mood would improve.	1	2	3	4	5
6	Eating [one or more specific foods] would feel wonderful.	1	2	3	4	5
7	If I ate something, I wouldn’t feel so sluggish and lethargic.	1	2	3	4	5
8	Satisfying my craving would make me feel less grouchy and irritable.	1	2	3	4	5
9	I would feel more alert if I could satisfy my craving.	1	2	3	4	5
10	If I had [one or more specific foods], I could not stop eating it.	1	2	3	4	5
11	My desire to eat [one or more specific foods] seems overpowering.	1	2	3	4	5
12	I know I’m going to keep thinking about [one or more specific foods] until I actually have it.	1	2	3	4	5
13	I am hungry.	1	2	3	4	5
14	If I ate right now, my stomach wouldn’t feel as empty.	1	2	3	4	5
15	I feel weak because of not eating.	1	2	3	4	5

Participant ID: **Date:** **Time:**

B.6 Sleep Diary

Daily Sleep Diary

Complete the diary each day ("Day 1" will be your first day). Don't worry too much about giving exact answers, an estimate will do.

Participant ID________________ The date of Day 1 ________________

	Enter the Weekday (Mon, Tues, Wed, etc):	Day 1	Day 2	Day 3	Day 4	Day 5
1	At what time did you go to bed last night?					
2	After settling down, how long did it take you to fall asleep?					
3	After falling asleep, about how many times did you wake up in the night?					
4	After falling asleep, for how long were you awake during the night in total?					
5	At what time did you finally wake up?					
6	At what time did you get up?					
7	How long did you spend in bed last night (from first getting in, to finally getting up)					
8	How would you rate the quality of your sleep last night? 1 2 3 4 5 (1=v poor, 5=v good)					

B.7 Post Study Document

Dear participant,

Thank you for your participation in the study so far. Once you have taken part in the post-study interview, you will have fully completed your participation – your efforts are greatly appreciated.

Now that the week of data collection is reaching an end, you will find instructions below on what to do with the measures after a full week of data has been collected. I suggest completing these tasks on Saturday morning, but if it is later than this rest assured that the data analysed will only be from Monday-Friday.

Log out and uninstall DeskTime
- Log out: Open the DeskTime interface. Click on your name icon (top right) and log out (do not delete your account as this will remove all of your data).
- Once logged out, you can uninstall the application from your device(s) (Windows – right click on DeskTime icon and select uninstall; Mac - locate application in Finder, then right click and move to trash).

Log out and uninstall RescueTime
- Open application on phone. Go to settings and log out (do not delete your account as this will remove all of your data).
- Once logged out, you can delete the application from your device.

Send questionnaire application data and delete app
- Use search function on phone to locate .csv file called AnalysisData. Email this file to [serena.midha@nottingham.ac.uk] as the file contains all of your app input data.
- Once you receive an email confirming the receipt of the file, you can delete the file and app.

Send evening questionnaires
- Complete the final sleep diary questionnaire and email it to [serena.midha@nottingham.ac.uk]. If you haven't already emailed the rest of your questionnaire responses, please do that also.

If you have any queries please don't hesitate to get in touch.

Appendix C

Study 3 Materials

C.1 Interview Questions Loose Guide

Personal Experiences

Before going into more specific questions, did anything in particular stand out to you about your mental workload experiences?

High mental workload:

- What did it feel like when you were experiencing a high mental workload level?
- Was the feeling the same for different instances of high mental workload? Or did the feeling of high mental workload depend on the task?
- *(If answer depend on task) Why?*
- Do you enjoy periods of high mental workload?
- *Why? When? Etc*
- Could you maintain a high mental workload level for a long time?
- *Does it depend on task? Why? Does anything affect your ability to maintain the level? Etc*
- When you're at a high mental workload level are you happy to stay at that level or do you seek out different levels?
- *(If seek change) When? Why? Would anything happen if you couldn't change level? How would you feel? Would your task be affected? etc for follow up*
- *(If happy to maintain) Why? What would happen if you were made to change? How would you feel? Would your task be affected? Etc*
- How would you feel if you realised that you may have a whole day of high mental workload levels ahead of you?
- *Would you take any actions? Etc*
- When you've had periods of high mental workload, would you say it affected any aspect of your behaviour or cognition or life?
- *Follow up eg if yes then what impact, methods to mitigate etc.*
- When you've had periods of high mental workload, do you feel as able to address life tasks? Such as the washing up.
- *Follow up why, how to manage, etc*
- Before we move on, are there any particular experiences of high mental workload that you would like to talk about?

Low mental workload and Medium mental workload:
Same set of questions asked for low and then medium mental workload levels.

We've been through your personal experiences of mental workload, and now we're going to look a bit deeper at how your experiences can be conceptualised.

Conceptualisation

Definitions:

- **There is actually no agreed upon definition of mental workload. Could you think of a definition that fits in with your experience of mental workload in daily life? If you need a starter, perhaps start with 'mental workload is….'**
- Even though there is no agreed upon definition, the components that mental workload is comprised of is generally agreed to be the cognitive and physical task demands (so the characteristics of the task) and the operator's experience of responding to the task, both of which can affect how well the operator is able to perform on the task. Could you think of a definition that fits in with these factors? If you are happy with your original definition you don't need to think of a different one.
- **Can you describe mental workload in general using a few key words?**
- **Can you describe what it feels like to experience high mental workload using a few key words?**
- **Can you describe what it feels like to experience low mental workload using a few key words?**
- **Can you describe what it feels like to experience a medium level of mental workload using a few key words?**

Metaphors:

- **If we become a bit more creative, can you think of a metaphor to describe mental workload as a concept? It can be as a whole, or it can consider different levels separately.** Just to remind you, a metaphor is a reference to one thing by mentioning another thing, e.g. a hairy dog could be described as a ball of fluff.
    - **If appropriate, draw it**
- Can you expand that metaphor to describe how it feels to experience low and medium mental workload levels?
- **Others have used metaphors which describe mental workload as: on a spectrum that changes throughout the day, walking up a mountain where the steepness of the walk reflects the mental workload level, a thermometer, a brain filled with a set number of bubbles and the number of bubbles that pop depend on the mental workload level, and an input-output relationship where you put in a certain amount of mental workload and expect to get a return. Do any of these resonate with how you would think about mental workload?**
- **Do you prefer your metaphor?**
- **Ok, so the final metaphor is….**
    - If CONTINUOUS – is described as continuous, as if mental workload is on a scale. Would you agree with that? Do you believe that mental workload levels don't have a limit, but any level can be used for any length of time? Do you believe that mental workload levels aren't categorised as low, medium and high, but are on more levels, like a scale?
    - If DISCRETE – is described as discrete, as if mental workload levels are separate and not related. Do you agree with that? Do you believe that mental workload levels have a limit, where different mental workload levels can be used a limited amount of times and for a limited

amount of time? Do you believe that mental workload levels are categorised into low, medium and high, instead of being on more of a scale level?

Colours and shapes:

- **What colour(s) and shape(s) would you associate with your experience of being at a high mental workload level?** There are different coloured pens here if it helps you. Why that colour?
- **What colour(s) and shape(s) would you associate with your experience of being at a low mental workload level?** Why did you make that choice?
- **What colour(s) and shape(s) would you associate with your experience of being at a medium mental workload level?** Can you explain your choice(s)?
- **What colour(s) and shape(s) would you associate with mental workload as a concept?** Can you explain why you chose that?

<u>Contextualisation</u>

Now we're going to look a bit deeper into the contextual side of tracking mental workload in daily life.

Hypothetical:

- Imagine that you owned an actual Fitbit for the brain – that is, a tracker that can objectively measure your mental workload levels in your life. You can reflect on that data by opening an app wherever and whenever you like.
- What would you use that data for? Would you like to track that data?

<u>Ethical considerations</u>

For the last part of the interview, let's talk about the ethical side of tracking cognitive activity.

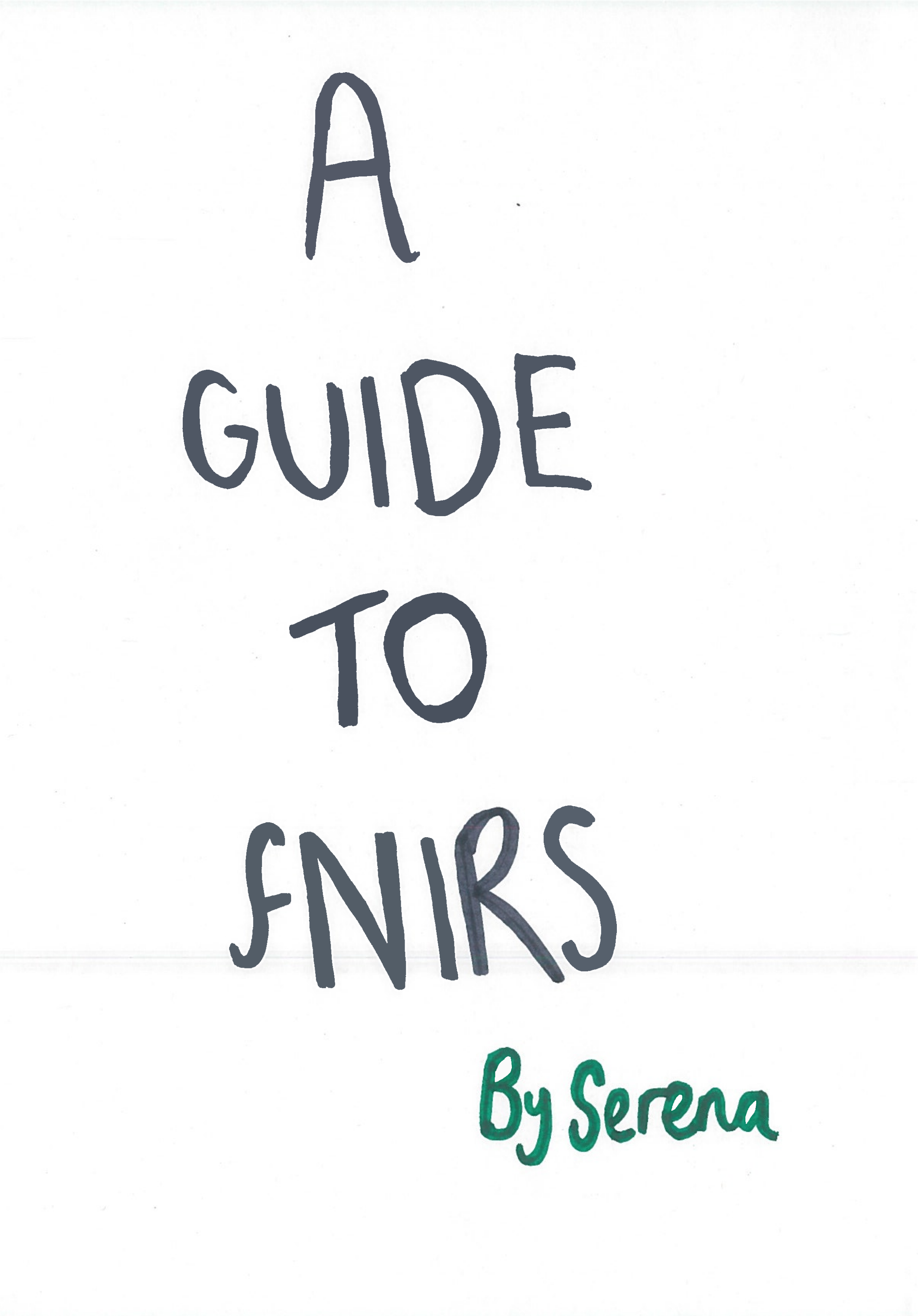
- If you were using the app for own personal improvement, would you have any concerns?
    - Do you have any concerns specifically about privacy?
    - Would these concerns be different to the concerns you may have about tracking physical activity data?
- What would your concerns be if this sort of data ended up in the wrong hands?
- Would you be prepared to share your data with people who might be interested in it, such as your boss?
    - Why?

- Imagine that you are a pilot and your boss wants to monitor your mental workload through this Fitbit for the brain. Their reasoning is that if your data is not optimal, you are putting other people at risk. What do you think about this?

Appendix D

Guide to fNIRS

D.1 Guide to fNIRS


A GUIDE TO fNIRS

By Serena

1

hemodynamic response

The ~~BOLD effect~~

~~- Blood oxygen level dependent~~

- Can be used as an indirect measure of neural activity
- Because blood CBF flow (increased oxygenation) is an indirect measure of neural activity
- This is neurovascular coupling - the coupling of blood flow with neural activity.

- Why CBF increases:
- Stimulus ⟶ increased neural activity ⟶ increased metabolic demand ⟶ decreased oxyhemoglobin (O_2Hb) and increased deoxyhemoglobin (HHb) in the area (which was already there) ⟶ increased blood flow to the area, because demand in oxygen, which rebounds with more oxygen than there was before the neural activity, so results in increased O_2Hb and decreased HHb in the end.
- So increased O_2Hb = increased neural activity
- decreased HHb = increased neural activity.

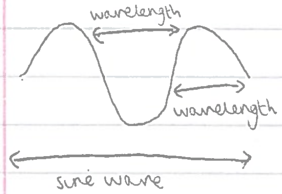
- Mental workload example:
- The n-back task is a working memory task
- The dorsolateral prefrontal cortex, parietal cortex, and anterior cingulate (among others) are areas ~~involved~~ involved in working memory
- When measuring the BOLD signal in these regions of interest there is increased neural activity with increased difficulty
- Increased metabolic demand because increased cognitive demand so likely increased number and firing rate of neurons, and maybe increased recruitment of brain areas.

2

Principles of fNIRS

LIGHT

- Electromagnetic field – a physical field produced by electrically charged objects.
- It can be viewed as the combination of an electric field (stationary charges) and a magnetic field (moving charges - currents)
- Electromagnetic radiation – the waves of the electromagnetic field
- Wave – oscillation (vibration that repeats the same pattern) that travels through space and time. Particles undergo vibratory motion about their mean position which repeats in equal intervals of time –
- Radiation is the emission of energy as electromagnetic waves (explains the name) or as moving subatomic particles
- Electromagnetic radiation waves of different frequencies/wavelengths are called different names because they have different sources and effects on matter – in order of increasing frequency and decreasing wavelength these are: radio waves, microwaves, infrared, visible light, x-rays, gamma rays.
- So visible light and infrared light are forms of electromagnetic radiation.
- Light exists in photons (positive energy charge) and each light wave has a wavelength and frequency.
- Electromagnetic Spectrum – the range of all possible electromagnetic radiation
- Visible light is the portion of the electromagnetic spectrum that is visible to the human eye – wavelengths between 400-700 nanometers (μm - one millionth of a metre
- The human eye sees each wavelength as a different colour – red has the longest wavelength (700nm) where infrared begins
- So infrared radiation is electromagnetic radiation with longer wavelengths (less frequency) than visible light – 700nm – 1mm.
- Infrared is invisible to the human eye.



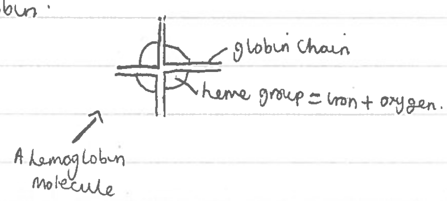
(wave is all the sine waves put together) (frequency is how often the event occurs)

TISSUE

- We are composed of 4 types of human body tissue
- This describes cells that are grouped together in a highly organised manner according to specific structure or function.
- These groups of cells form ~~tissues~~ tissues, which then make up organs and various parts of the body
- Muscle tissue - made up of long and fibrous cells ready for contraction
- Epithelial tissue - made up of epithelial cells which are flat, like a cube, or upright and they are joined tightly together like a tightly stitched quilt. Designed for protection, epithelial tissue makes up ~~sk~~ skin.
- Connective tissue - makes a connective web that holds body parts together and provides support. Does this by filling the spaces inside the body with fibres within a liquid, solid or jelly-like substance
- Nervous tissue - found in the nervous (neurons transmitting signals) system and make up neurons

HEMOGLOBIN

- Because oxygen is not very soluble (dissolvable) in water, we need an efficient oxygen-carrying molecule in order to get oxygen around the body.
- These ~~m~~ molecules are the proteins hemoglobin and myoglobin
- Hemoglobin is contained in red blood cells and is the oxygen carrier in blood.
- A hemoglobin molecule is made up of 2 subunits.
- These are a folded polypeptide chain ~~s~~ - the globin chain. Polypeptides are chains of amino acids. Amino acids are compounds that combine to form proteins. So the globin chain is a protein
- And a heme group - a type of porphyrin. A porphyrin is not a protein - it is formed from carbon and nitrogen. ~~Each~~ Hemes are a compound that contains iron. There are 4 heme groups in a hemoglobin molecule.
- The oxygen-binding ~~space~~ site of hemoglobin is the heme pocket. ~~A single~~ A single atom of oxygen forms a reversible bond with the iron in the heme groups, so a hemoglobin molecule binds 4 oxygen molecules - oxyhemoglobin.



MEASURING

- Human tissue has relatively low absorption of red and infrared light (if you place your fingers over a white light they will appear to glow red because the finger tissues have absorbed all the visible lights except red which can pass through the tissue - it's quite transparent).
- Hemoglobin (oxy and deoxy) can absorb ~~some~~ infrared light.
- Oxyhemoglobin and deoxyhemoglobin have different absorption spectra (they absorb different ~~amounts~~ amounts of light at different wavelengths) in the infrared range - oxyhemoglobin absorbs less infrared light compared to deoxyhemoglobin at lower wavelengths in this portion of the electromagnetic spectrum, whilst oxyhemoglobin absorbs more light than deoxyhemoglobin at higher wavelengths of the infrared range
- There is an optical window between 700nm - 900nm where the light is only absorbed by hemoglobin (oxy and deoxy) without interference from other factors and can penetrate deeply. Beyond this wavelength, most of the photons (infrared light photons) are absorbed by water.
- To determine the (amount of) concentration of oxy- and deoxyhemoglobin, a light source (emitting optode - an optode is an optical sensor) emits ~~[illegible]~~ infrared light at the scalp
- A distance away, a photo detector (recieving optode) detects the intensity of the reflected light
- The intensity of the reemerging reflected light will have changed from the intensity that was emitted from the light source because of the absorption of light from the hemoglobin and scattering.
- For example, if there is an increase of hemoglobin, more photons will be absorbed and scattered so the intensity of the reflected light will be reduced.
- Therefore, changes in light intensity can determine hemoglobin concentration.
- This can be calculated using the modified Beer-Lambert law. in this case - from light source to photo detector
- (Scattering is where the light photons are forced to deviate from a straight path (trajectory) due to non-uniformities in the medium, such as ~~[illegible]~~ fluid fluctuations)
- (2 wavelengths of infrared light are emitted from the light source to accont for the different absorption spectra of oxy and deoxy, so ~~both~~ concentrations of both oxyhemoglobin and deoxyhemoglobin can be calculated)

(The main interaction between the emitted light and the tissue is absorption, but scattering accounts for the photon loss measured in the reflected light that is not caused by absorption - so false conclusions about hemoglobin concentration might be made if scattering is not accounted for.) - Scattering also means that photons will reach the detector but take a longer path

MODIFIED BEER-LAMBERT LAW

- (measure of hemoglobin concentration)

optical attenuation (reflected light intensity) (reduction in intensity)

incident (emitted) light intensity (at that certain wavelength)

d is the distance between the source and detector

Differential pathlength factor - indicates the mean of how many times ~~foo~~ further than d the detected light has travelled (at that certain wavelength)

$$I(\lambda) = I_0(\lambda)e^{-\mu_a(\lambda)\, d\, DPF(\lambda)} + G(\lambda)$$

wavelength

So reflected light intensity at ~~that~~ a certain wavelength (either for oxy or deoxy remember)

absorption coefficient of the hemoglobin at that certain wavelength (a constant that is different for oxy and deoxy)

A geometry dependent constant to do with photon loss due to scattering

(at that wavelength)

(Repeat modified beer-Lambert ~~law~~ law for ~~a~~ the two oxy and deoxyhemoglobin wavelengths)
(simultaneous equations for oxy and deoxy can be solved using matrix inversion)

(so [illegible] the more brain can be measured / increase of channels = ~~we~~ measure more parts of the brain)

(A channel is the tissue at the midpoint between the source and detector and is at a depth of around half of the source-detector separation)

SOURCE-DETECTOR DISTANCE

- Increasing the distance between the source and detector of near-infrared light will generally increase the amount of photons that are detected from deeper tissues (deeper brain layers) in the brain, meaning the signal (detected light) should have an increased sensitivity to the oxygen levels in the brain.
- However, even though there would be the ability to detect ~~signal~~ photons from deeper parts of the brain, fewer photons in general (from the source) will ~~be~~ reach the detector because of scattering, meaning that the measured signal (light intensity) will have a lower signal-to-noise ratio (because signal is lost with the scattered lost photons so the measured signal is noisy)
- So the larger the source detector distance, the more brain can be measured, but the lower the signal-to-noise ratio.
- A source-detector distance of 30mm is a good compromise between having reasonable sensitivity (coverage) ~~of~~ to the brain and having enough detected light to have an acceptable signal-to-noise ratio. (Artinis Octamon has a 35mm source-detector distance).

- Another problem with interpreting the signal (light intensity) is the contamination from the scalp and skull tissue
- The human scalp and skull tissues are highly vascularized (contain blood vessels (oxygen, hemoglobin)) so they are thick (13mm thick in total)
- Light from the source to the detector has to pass through these layers twice (on the way into the brain and out of it)
- The fNIRS measurements are very sensitive to the scalp and skull tissues, meaning that the signal contains the effect of hemoglobin concentration changes in the scalp and skull as well as in the brain. So the signal is contaminated ~~and~~ as it does not just contain the ~~effects~~ hemoglobin concentration information from the brain.
- ~~To remove~~ Removing this contamination is a challenge for fNIRS technology.
- Hemoglobin changes in the scalp and skull arise from cardiac activity, respiration, ~~changes~~ changes in blood pressure, and vasomotion, which can be correlated with hemodynamic response
- It is important to use effective methods to remove these noise confounds in order to determine the true brain oxygenation levels and avoid mistaking superficial signals for functional activation.
- A promising solution to this problem are short-separation channels

- By using a channel with a very small source-detector separation distance, the depth sensitivity can be exploited
- Because the further apart the source and detector, the deeper the light photons can penetrate into the brain, remember
- So by using a very short separation distance, the photons can't penetrate deeply.
- At the right correct short distance, ~~only~~ it is possible to only probe the extra-cerebral tissues (the scalp and skull), so hemoglobin concentrations in only those areas (without any signal from the brain) can be measured.
- The signal measured by this short separation channel can then be used to regress the superficial noise from the standard fNIRS signal.
- So the functional brain response can be isolated without the contamination from the scalp and skull hemoglobin concentration
- It has been suggested that 8.4mm is the optimum ~~sort~~ short-separation channel distance for human adults
- The short separation channel should be located no more than 1.5cm from the ~~long~~ standard channel in order to effectively regress out the superficial layer noise. This is because the systemic physiology ~~of the body tends to~~ is not homogeneous across the scalp.
- So there should be multiple short separation channels

TYPES OF fNIRS

- Continuous wave -

- Most commercially available systems use continuous wave
- This is where the sources emit near-infrared light at ~~a~~ (usually) two wavelengths and changes in light intensity can be related to changes in relative hemoglobin concentration ~~and~~ calculated by using the modified Beer-Lambert law.
- (Absolute concentration/change is the actual number, calculated by a subtraction). (Relative concentration/change is a ratio (the quantitative relation between 2 things - eg there are 8 oranges and 4 apples = a ratio of 8:4, or 2:1) - ~~so~~ relative change in hemoglobin is the difference in ratio between two periods of time). (Eg, a company employed 20 employees last year and 10 employees this year - the absolute change is -10 but the relative change is -50%.)
- The measure is relative because the baseline (beginning) levels of hemoglobin cannot be calculated so the measures start from ~~zero~~ zero ~~baseline~~ and the concentration changes of hemoglobin are calculated relatively to that baseline.
- The absolute quantification of hemoglobin levels cannot be calculated because continuous wave methods do not know the ~~exact photon~~ path length of the photons due to scattering.

- Frequency-domain -

- Measures absolute hemoglobin concentration
- Works by using laser sources to emit near-infrared light at 100MHz (remember flashing balls at rate 9.0Hz). Then calculates the back-scattered amplitude which ~~can~~ provides the information about scattering, so absolute measures of hemoglobin can be calculated
- More technically complex than continuous wave systems.

- Time-domain -

- Also measures absolute hemoglobin concentrations
- A light source emits a few picosecond (a trillionth of a second) pulses. A fast time-resolved detector measures ~~recovers~~ the time of flight (~~aka~~ referred to as the temporal spread function). The temporal spread function provides information of the ~~absorbed~~ scattered light and the depth reached by the photons in the brain
- The most technically complex method.

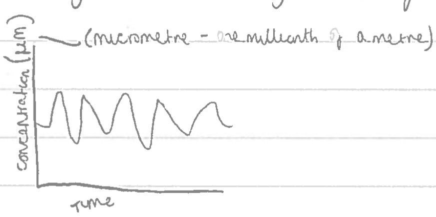
3

~~Analysis~~ Data pre-processing

~~FILTERS~~

THE SIGNAL

- Oscillations are movements back and forth
- fNIRS shows the hemodynamic response (BOLD signal is easier to think about) oscillations
- So the signal shows every moment of the BOLD signal (hemoglobin concentration) – the signal fluctuates up and down

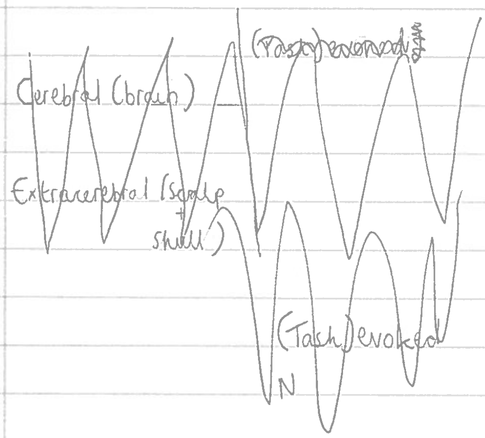


NOISE

- The signal is contaminated by noise – signals from other sources that interfere with the interpretation of the ~~[illegible]~~ signal we are interested in.
- As we want to know the hemodynamic response, the noise needs to be removed/corrected for in order to isolate the hemodynamic response and increase the accuracy of our interpretations.
- There are 3 sources of noise that can characterise fNIRS measurements:
- Instrumental noise: is electrical noise from the computer or ~~[illegible]~~ other hardware. Much of this noise is high frequency and can be filtered out using a low-pass filter. This noise is nearly negligible compared with the other sources of noise, particularly physiological noise
- Physiological noise: The largest source of noise – already discussed under source-detector separation and next page – this noise is (mainly) "the physiological noise in the superficial layers" ~~[illegible]~~
- Motion artifacts (under experimental errors): caused when movement of the head causes motion between the optical fibers and the scalp. This ~~motion causes an rapid shift~~ changes the coupling between the optical fibers and the scalp which results in a period of high-frequency noise/spike and/or a shift in the baseline measurements of oxyhemoglobin and deoxyhemoglobin (the baseline is the concentration (usually value) that the fash or other recorded data is compared against – so this baseline concentration is likely to change (increase or decrease) if there is a re-coupling so the other recorded data after this shouldn't be compared to concentrations from the old coupling).

- Movement artifacts are larger and occur more rapidly than genuine hemodynamic changes and are more irregular than systemic physiological noise.
- The spikes are usually ~~characterised by high frequen~~ rapidly occuring and are much larger in magnitude than the ~~genuine~~ functional hemodynamic response. ~~The~~ Algorithms to detect motion rely on these characteristics to distinguish these artifacts from other signal changes.
- Manually, motion artifacts are relatively easy to identify in the fNIRS signal by combining observations of the subject during the fNIRS recording and then visually inspecting the data.

- A closer look at what EVERY fNIRS signal is composed of (does not contain motion artifacts because they are not always present):



	(Task) evoked	
	Neuronal (neuronal activity)	Systemic (from a system (system noise) – (hemoglobin) increase in blood flow
Cerebral (in the brain)	Functional brain activity – neurovascular coupling (what we want to isolate)	eg, changes in blood pressure, carbon dioxide
Extra-cerebral (scalp + skull)	— Nothing	eg," changes in blood pressure + carbon dioxide

already discussed in short-separation section but remember that this also happens in the brain as well as scalp and skull

	Non-(task) evoked – will be happening at all times	
	Neuronal (neuronal activity)	Systemic (increase in blood (hemoglobin) flow from a system)
Cerebral (in the brain)	Spontaneous brain activity – neurovascular coupling	e.g., heart rate, respiration, meyer waves (low frequency waves from blood pressure) and very low frequency waves
Extra-cerebral (scalp + skull)	– Nothing (no neurons in scalp and skull?)	" "

~~NOISE CORRECTION~~

- The different types of noise (mainly referring to physiological and motion) need to be addressed individually, through ~~on~~ certain filters methods can be ~~applied~~ used to correct for physiological and motion artifact noise.
- Combining different correction methods ~~allows~~ / filters allows us to take advantage of different types of correction algorithms (different ~~method~~ filters address different aspects of the signal so increase artifact removal).

PHYSIOLOGICAL NOISE CORRECTION

- AFTER the fNIRS signal has been checked and corrected for MOTION artifacts, the physiological noise needs to be addressed.
- Short separation channels seem to be the gold-standard approach for ~~to~~ correcting extra-cerebral (superficial layer) noise. But this needs to be used in combination with a filter that can correct for cerebral (mostly systemic) noise. (a big problem (source of contamination) in the fNIRS signal)

BANDPASS FILTERS

- Removes ~~specific~~ frequencies outside of a specific range
- Consists of a low-pass (removes high frequency data in the signal) and a high-pass (removes low frequency data) filter
- The low-pass filter can remove noise such as fast cardiac oscillations and the high-pass filter can be used to remove noise such as heartbeat or mayer waves.
- Types of bandpass filter include Butterworth filters, elliptic filters and Chebyshev filters, but there is currently no reason to favour one of them. The Butterworth filter is most commonly seen in the literature.
- It is recommended to use a bandpass filter with cut-off frequencies between 0.01Hz (high-pass) – 0.5Hz (low-pass) (so frequencies (oscillations) outside of this range are removed from the signal). The 0.5Hz low-pass frequency is quite high/extreme in order to avoid accidental removal of functional hemodynamic response (actual brain activity). Thus, a bandpass filter could be used in conjunction with more sophisticated/advanced filter methods to correct for physiological noise
- Can be an effective method for removing many sources of physiological noise in the data (probably should always use this filter)
- But it cannot remove physiological noise that ~~has~~ is in the same frequency range as the functional hemodynamic response, such as respiration ~~of~~ signals (because the respiration noise is not outside the brain activity frequency)

ADAPTIVE FILTERS

- Another method of removing physiological noise
- Adaptive filters do not consist of constant filter coefficients (parameters) and previous information about the signal and noise does not need to be known.
- This is unlike fixed filters which require information about the signal and noise before applying a set of fixed rules to remove noise.
- So filters with adjustable parameters are known as adaptive filters
- Adaptive filters need two inputs - a primary input containing the corrupted signal and a reference input containing just the noise that is correlated with the noise in the primary signal (in fNIRS terms this is the short separation channel)
- Adaptive filters are defined by four aspects:
- 1: The signals (primary and reference) being processed by the filter
- 2: The structure that defines how the output signal is computed from its input signal
- 3: The parameters within this structure that can be iteratively (repeatedly) changed
- 4: The adaptive algorithm that determines how the parameters are adjusted from one time instant to the next
- How the adaptive filter works: the primary input is the corrupted signal. The reference input is the noise which is uncorrelated with the signal in the primary input but is correlated with the noise in the primary input. The reference signal passes through an adaptive filter to produce an output that is a close estimate of the noise in the primary signal. This noise estimate is subtracted from the corrupted ~~signal~~ primary signal to produce an estimate of the true signal.
- ~~Adaptively~~ The parameters need to self adjust in order to minimise the error signal (think about it as the filter looks at the signals moment by moment and adjusts to remove different levels of noise throughout the primary signal instead of treating it as a fixed and consistent amount of noise that is removed using the same numbers throughout the signal)
- Types of adaptive filter (structure) (the filter equation that computes the output from the input): Finite Impulse Response (FIR) and Infinite Impulse Response (IIR) -
- These filters differ mathematically as IIR uses some of the filter output as input which makes it a 'recursive' function (repeatedly applying a given relation used to predict the properties of the noise). IIR is faster to compute because it requires less numbers of terms but it has stability issues. FIR is very stable and reliable and seems to be most commonly used.

- Types of adaptive algorithm (that determines how the parameters are adjusted moment to ~~moment~~ moment): Least Mean Squares (~~n~~LMS), ~~NLM~~ Normalised Least Mean Squares (NLMS), Recursive Least Squares (RLS).
- RLS is the 'ultimate' adaptive algorithm but it has implementation issues. LMS and NLMS have a good reputation and are most commonly used. (Zhang 2012 developed good RLS method)
- E.g. 'The adaptive filter used a FIR.... The Widrow-Hoff LMS adaptation algorithm was used....' (Zhang).

PRINCIPAL COMPONENT ANALYSIS (PCA)

- Is another method of removing physiological noise
- It is a dimension-reduction tool (filter) that can be used to reduce a large set of variables (in this case the variables are the components that make up the signal) to a small set of new variables that still contains most of the information in the large set
- These new variables are called principal components and they account for the majority of the variability in the data. This enables us to describe the information (signal) with considerably fewer variables (components) than was originally present.
- The first principal component ~~accounts for~~ explains the most variability in the data, and the second ~~each~~ succeeding ~~pc~~ principal component accounts for ~~as much of the~~ the maximum amount of the remaining variability, and this continues for the other principal components
- The principal components are combinations of the ~~origi~~ original variables. As they are ordered by how well they predict the dependent variable (the whole signal), we know which variables are most and least important/influential on the signal.
- The principal components created from the PCA are all uncorrelated with each other
- In fNIRS the physiological noise is covariant (covariance is how changes in one variable are associated with changes in another variable - an unstandardized correlation) ~~across~~ among the different channels (as noise increases in one channel it will increase in the other channels). So by eliminating the ~~variable~~ principal component with the physiological noise the signal to noise ratio can be increased
- The maths algorithm for calculating the filter is complex and we do not need to understand ~~this~~ it, but we can understand a simple outline of what the algorithm does:
- A matrix is calculated that ~~su~~ describes how the variables relate ~~related~~ to one another.
- The matrix is then broken down into two separate factors:

(picture it plotted on a graph)
- 1- the directions in which the data is dispersed (called eigenvectors)
- 2- the magnitude (importance / contribution) of these different eigenvectors (eigenvalues)
- The original data is transformed to align with the different directions, so new variables (principal components) ~~[illegible]~~ are created according to which direction the data is dispersed and are combinations of the original variables.
- So PCA combines correlated predictors and allows us to drop the eigenvector containing physiological noise
- PCA takes advantage of the spatial ~~[illegible]~~ diversity, ~~between~~ which is the difference in spatial behaviour between the evoked signal and the physiological interference. Usually, filters are in the temporal domain.
- However, PCA algorithms require the physiological noise eigenvector to be calculated from a separate baseline data set rather ~~[illegible]~~ than from the functional data. Then these spatial patterns from the baseline condition can be used to remove similar covariance patterns in the functional data.
- PCA methods have also been questioned because they rely on the assumption that single phenomena lead to single eigenvectors (so assumes that physiological interference signals will follow the same direction).
- PCA is also not suitable when there are a small number of channels
- PCA is viewed as preferable compared to independent component analysis (ICA)

EXTERNAL MEASURES
- Another approach to removing physiological ~~or~~ noise
- This method involves physically measuring sources of physiological noise in parallel to measuring the fNIRS signal
- E.g measures such as blood pressure, skin conductance and heart rate.
- This means that all the components of the fNIRS signal can be quantified and the sources of physiological noise can be removed from the signal.
- If done properly, all of the noise ~~[illegible]~~ can be identified and therefore removed, but this requires many external measurements.

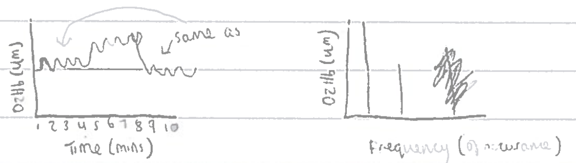
MOTION ARTIFACT CORRECTION
- ~~[illegible]~~ Even though motion artifacts can be visually identified relatively easily, there are ~~[illegible]~~ ~~[illegible]~~ algorithms that can automatically detect motion artifacts by identifying changes

in the fNIRS signals which are of a scale, rate or nature that are unlikely to be physiological
- However, once motion artifacts have been identified there is no well-established approach to their removal or correction
- A common approach is to have enough stimulus trials to minimise the average impact of the artifacts and even more common is to reject all trials where a motion artifact has been detected
- However, this ~~only approp~~ approach is only suitable if the number of detected motion artifacts are low and the number of trials is high, otherwise there will not be enough data to investigate the true hemodynamic response.
- So in these cases motion artifacts need to be ~~corrected for~~ reduced or removed whilst preserving the rest of the data.
- Complimentary measures have been proposed to correct ~~[illegible]~~ motion artifacts. ~~[illegible]~~ These work ~~on~~ as a reference that measures the motion artifact signal but is not correlated with the hemodynamic response. Such as an external measure of motion from an accelerometer which can be used to regress the motion artifact from the standard fNIRS signal. Or by using co-located channels which means having ~~a source and~~ all the sources (more than usual) and detectors in ~~the~~ exactly the same location (this works because the co-located channels ~~can detect~~ are sensitive to motion but not the ~~fNIRS signal~~ hemodynamic response so the motion can be filtered out of the fNIRS signal)
- These approaches are promising
- Other approaches for correcting motion artifacts are post-processing techniques - filters to correct for motion after data collection without complementary measures that can alter the experimental paradigm.
- 2 studies have compared the use of different motion artifact correction filters.
- Including PCA, spline interpolation, wavelet analyses, Kalman filtering and correlation-based signal improvement
- Both showed that it was better for the signal-to-noise ratio to always correct for motion than leave it or reject trials ~~[illegible]~~
- Both found wavelet analyses to be ~~the most~~ a very effective filter for correcting motion artifact
- One found spline interpolation also very effective
- ~~Spline interpolation~~ This is because spline interpolation should only be used/is most effective when motion artifacts are easily detected, so it depends on the data set.
- It is suggested that wavelet filtering has the potential to become the standard motion correction technique

- Motion artifacts are variable by nature - they can ~~have~~ be high frequency, high amplitude spikes which are easily detectable or they can have lower frequency content which is harder to distinguish from the hemodynamic response. They can ~~gececbly~~ generally be classified into three categories - spikes, baseline shifts and low frequency variations, so the motion artifact correction technique can depend on the data set.

WAVELET ANALYSIS

- Signals are often recorded in the time domain - how the amplitude (how height of the oscillation / y axis values) changes over time (so time on x axis)
- But signals ~~are~~ often ~~are~~ provide more useful information in the frequency domain - the frequency of the amplitude over that time period (frequency on x axis, amplitude on y axis).
- For example, ~~they~~ an fNIRS signal will record levels of (e.g.) oxyhemoglobin over time:



- But we might want to deal with the signal ~~on~~ according to how often each ~~level~~ oxyhemoglobin level occurs.
- The process of getting from the time domain to the frequency domain and from the frequency domain to the time domain is called the Fourier Transform. This works because any signal can be represented by an equation which adds up sin() and cos(). Sin() and cos() waves are continuous oscillations at a constant amplitude (can be different frequencies). Adding these waves continues until infinity. Any action taken on the signal, such as reducing noise, requires manipulation in the frequency domain so the Fourier Transform is done before these processes. The signal can then be transformed back to the time domain - which might look different to the original signal in the desired way.
- However, a limitation of the Fourier Transform is in regards to Heisenberg's ~~un~~ Uncertainty Principle. In physics this basically means that you can know where a particle is or how fast it is going, but not both. ~~[illegible]~~ Or in other words, if you want to become more certain about the position of the ball, you have to become less certain ~~about~~ of the speed of the ball and vice versa (the amount of uncertainty is called resolution).
- The Fourier Transform has this problem - you can either be sure of the frequency or time of the signal, but not both.
- The problem is that when you are handling a real signal it would be useful to know the frequency of the signal at an exact moment in time.

- This is where wavelets are useful
- Wavelets are mini waves - unlike sine and cosine waves which go on forever, wavelets are a short burst of waves that fade quickly to zero (e.g. one form of a wavelet -)
- As sine and cosine waves are infinite, it is not possible, for example, to know when in time an exact frequency occured, because each frequency exists across all time
- The Wavelet Transform deconstructs (breaks down) the signal into loads of wavelets being added together
- The fact that wavelets are limited in time and frequency means that there is more resolution (less uncertainty) about the time when looking at frequency. This is because we can 'slide' the mini wave (wavelet) along the x axis
- This is done in the wavelet domain and not frequency domain. So time is along the x axis and frequency is represented by scales. In sine and cosine waves, high frequencies squash the wave and low frequencies are stretched . This also happens to wavelets which represent frequencies at different scales. E.g. a high frequency scale is and low frequency scale
- So as the waves are small it is possible to have higher frequency and time resolution information
- In fNIRS, wavelet Analyses can be an effective method of removing motion artifacts because as the frequency and time of the signal is known, abrupt changes in the hemodynamic signal that characterise spike (the majority) motion artifacts can be isolated
- When using this filter for fNIRS motion artifacts, it essentially works by:
- The time-domain signal is passed through high-pass and low-pass filters to filter out portions of the signal. This is repeated (every time it's done again it's counted as a new level) E.g. if our signal has frequencies up to 1000 Hz, at the first stage we split the signal into two parts by passing it through a high-pass and low-pass filter (so we have the low-pass portion corresponding to 0 - 500 Hz and high-pass to 500 - 1000 Hz). We repeat this which creates more portions. We continue until we have decomposed the signal to a pre-defined level. So we've ended up with a bunch of signals (wavelets) that add up to the original signal but each of them represent a different frequency band. So we know which frequency happened at which time.

(just a coefficient generated from the high-passed filter data at that level)

- At each level, a detail coefficient (from the high-pass filter) and an approximation coefficient (from low-pass) is obtained. The model assumes that the measured signal is a linear combination of the hemodynamic response and the motion artifacts, that the detail coefficients have a Gaussian (normal) distribution (the high frequencies have a normal distribution(?)), and the hemodynamic

response ~~has~~ is smoother and slower than motion artifacts. Thus, it is expected that the detail coefficient accounting for the evoked hemodynamic response will be centred around zero with low variance, while the outliers of the normal distribution are the detail coefficients accounting for the motion artifacts. These outlying coefficients are set to zero before reconstructing the signal from the wavelet domain back to the time domain ~~&~~ should remove the motion artifacts.
It is suggested to set the outlying coefficient parameter to 0.1.
- This has described the discrete wavelet Transform which is highly regarded and commonly used. Other methods exist such as the continuous wavelet Transform.
- Wavelet analysis works for fNIRS data because the signal is non-stationary - the signal changes over time (if it was stationary all frequency components would exist at all times)
NB
- Basically, Wavelet Analyses is based on the assumption that motion artifacts have different characteristics in terms of amplitude and duration from the original signal. This difference is better highlighted in the wavelet domain due to ~~a~~ good localisation of time and frequency.

SPLINE INTERPOLATION
- Interpolation is a method of estimating values (or creating new values) between ~~around~~ known values (e.g. if we sampled at timepoints 1, 2, 3, 4, 5, 6 but want to know what the value of the signal was at timepoint 2.5 we could use interpolation. Likewise if a motion artifact is detected we could use interpolation to estimate ~~the~~ how the signal should have been).
- A simple interpolation ~~m~~ method is linear interpolation - this ~~a~~ is an equation that basically joins all the known values up on a graph and the estimated value is the point at which the line intercep[ts] the ~~value~~ timepoint (e.g. 2.5). This isn't a very precise method.
- Polynomial interpolation is a better method. A polynomial is 'many terms' (e.g. $4xy^2 + 3x - 5$ is 3 terms) that can be combined through addition, subtraction, multiplication and division. A polynomial can have terms that are constants (e.g. 5), variables (e.g. the x in $5x$) and exponents (e.g. the 5 in y^5). The degree of a polynomial is the largest exponent of a variable in the polynomial (e.g. ~~an~~ $5x^6 + 2x^2 - 4$ ~~therefore~~ this is Degree ~~of~~ 6). Polynomial interpolat[ion] is a polynomial ~~with~~ which goes through each known data point linking them up in a curve. The estimated value (e.g. for timepoint 2.5) is calculated from the variables in the polynomial.
~~an~~ However, polynomial interpolation ~~is~~ sometimes produces a curve that doesn't look right (is ~~no~~ contrary to common sense).

- Spline interpolation is a better method. Spline interpolation uses a number of polynomial (~~one of~~ one for each known data point) and adds them together to ~~make~~ link up the data points in a smooth curve. Again the ~~max~~ estimated value is calculated from the variable in the polynomials. Spline interpolation has low-degree polynomials compared to polynomial interpolation. A cubic spline interpolation means that the polynomials have a degree of 3.
- In fNIRS, to correct for motion artifacts a channel-by-channel approach is applied. It only acts ~~off~~ on motion artifacts and leaves the rest of the signal unmodified. Motion artifacts firstly need to be identified in each channel using an automatic detection algorithm (Homer2 function, not related to spline interpolation). The period of motion artifacts are then ~~modelled~~ modeled by cubic spline interpolation. This result is then subtracted from the original signal (which had ~~the~~ the motion artifacts) to correct for the motion artifacts. To ~~m~~ make the signal a continuous signal after the spline interpolation, the ~~motion~~ segments where the motion artifacts were are shifted by a value calculated from the mean value of that segment and the mean value of ~~the~~ a segment before that segment. Spline interpolation needs a parameter to be input ~~to~~ to determine the degree of the spline function. 0.99 is the recommended value for a cubic spline interpolation.
- Spline interpolation relies on a reliable technique to detect motion artifacts before spline interpolation is implemented. If the artifacts are difficult to detect (not been detected by the technique), spline interpolation will not be applied properly and thus the signal will not be improved as a number of motion artifacts will not have been corrected for.
- So seems like an effective method when motion artifacts ~~can be~~ are detectable.
- Spline interpolation also has the ability to ~~correct~~ remove baseline shifts (so does PCA) - so if there are sudden shifts in the data (baseline shifts) this method can be used to remove them - ~~A~~ this might mean that spline interpolation is used to correct the baseline shift if present (might not be present) and then a different method is used to correct for the ~~spikes~~ spikes (like wavelet) if spline interpolation is not appropriate for that dataset (hybrid methods (suggested by Yücel)).

4

Data analysis

BASELINES

- Remember that continuous wave fNIRS measurements are relative - the changes of blood oxygenation concentration relative to another time point (the baseline) are measured as the absolute values of blood oxygenation concentration can't be measured.
- A baseline is a 'resting state' so can be taken as the ~~immediate~~ immediate seconds before ~~or~~ a stimulus or a baseline condition (for e.g. 1 minute) can be recorded.
- So as continuous wave fNIRS can only measure a change in hemoglobin concentration (expressed in μm) and cannot provide information about the absolute starting value, the baseline values are arbitary (don't mean anything) and are different for different people/days/times
- Therefore, if two people have the same amount of increase in oxyhemoglobin, the measures will be on different scales because they have different arbitary starting values.
- Baseline correction should be used in order to compare the data among participants by putting all the data on the same scale
- This is done by taking the average value of the baseline and subtracting it from all the other (task data) data points collected (so baseline set to arbitary value of 0 ~~and all the data~~ which is the same for all participants as well as the relative changes in blood oxygenation ~~concentrat~~ concentration).

- E.g. -

	(average) Baseline	Condition 1	C2	C3
Participant 1:	1	3	4	5
Participant 2:	3	5	6	7

(just for this example the baseline average value was subtracted from one ~~mean~~ value from the condition - in reality the average baseline value is subtracted from all the data points collected after the baseline)

=

	Baseline	C1	C2	C3
P1 :	0	2	3	4
P2 :	0	2	3	4

now on same scale

- So baseline correction accounts for the (arbitary) individual variability of fNIRS data
- Baseline normalisation can also do this

AVERAGING

- Depending on the study design, ~~averaging~~ calculating an average (hemoglobin concentration change from baseline) of each condition for each channel, and comparing the averages, might be appropriate (e.g. block design) (well, using statistics based on averages)
- However, when comparing averages information about the exact shape and timing of the hemodynamic response is lost.

- Combining data from a number of ~~[illegible]~~ channels across a region of interest can also be appropriate (e.g. computing an (one) average ~~values~~ value for channels 1-4 for each condition and comparing the values).

~~[illegible]~~

THE GENERAL LINEAR MODEL (GLM)

- Analysis method that overcomes the issue of averaging as it considers the entire signal time course which provides more statistical power. The GLM is the standard analysis approach for fMRI which has now been extended ~~for~~ as an analysis method for fNIRS data as both imaging methods are based on the hemodynamic response
- Maybe start by reading Andy Field's explanation of regression - this is a regression ~~technique~~ approach.
- The model is:

$$\boxed{y = x\beta + e}$$

- The GLM describes the ~~BOLD~~ BOLD response in a single voxel (or in our case the GLM describes the hemodynamic response in a single channel) (y) in terms of all of its contributing factors ($x\beta$) in a linear combination, whilst also accounting for the contribution of error (e).
- A simple way to think of it - we start with a baked cake (y) and try to explain how it is made. We specify which ingredients were included (x). For each ingredient, the GLM finds the ~~quantity~~ quantity (β) that produces the best reproduction of the baked cake. Then if we try to bake the cake with what we know about x and β then the error (e) would be the difference between the original cake and ours. So the model for the cake would be:

$$\overset{\text{cake}}{y} = \overset{\text{flour}}{x_1\beta_1} + \overset{\text{sugar}}{x_2\beta_2} + \overset{\text{butter}}{x_3\beta_3} + \overset{\text{eggs}}{x_4\beta_4} + e.$$

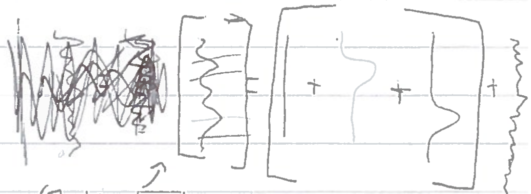
- So y is the dependent variable (hemodynamic response) - at each time point so we have the whole.
- x is the independent variable(s) (our predictors of the hemodynamic response - eg. the experimental time course of it (the hemodynamic response at each time point - so there are many y's) (

$\boxed{y = x\beta + e}$

conditions. x embodies all available knowledge about experimentally controlled factors and potential confounds). For each x (predictor), its time course (its predicted hemodynamic response for the whole measured signal time (y)) is obtained by convolution of a box-car function that reflects the experimental design with a hemodynamic response function. This means -

- A box-car function is a short off-on-off function that reflects what the neuronal response might be - e.g. for a short stimulus it might look like and for long stimulus
- from the box-car function of our predictor (x) the hemodynamic response function can be predicted (the hemodynamic response function is a predicted hemodynamic response for one single presentation of the stimuli) e.g.
- If several of the predictor stimuli are presented the box-car function might be and the predicted hemodynamic response might be . But if several stimuli are presented in different time intervals like then the hemodynamic response might not have time to return to baseline so it might be predicted to look like
- The prediction of the hemodynamic response from the box-car function that reflects the stimulus and the hemodynamic response function is convolution.
- β are the model parameters (beta weights) (the slope of the line). Each x (predictor) gets a β value which quantifies its contribution to the measured time course y. Each x value (there are several which reflects the whole time course of y) is multiplied with the β value that is associated with that x (so one beta weight for each predictor - not one beta weight for each value of one predictor) and these values of all the xs at each time point are added together to get y.
- Because there are many y values (the hemodynamic response over the whole time course) the GLM equation actually represents a matrix:

$$\begin{bmatrix} y \\ y \\ y \\ y \end{bmatrix} = \underbrace{\begin{bmatrix} x_1 & x_2 & \cdot & \cdot & x_n \\ x_1 & x_2 & \cdot & \cdot & x_n \\ x_1 & x_2 & \cdot & \cdot & x_n \\ x_1 & x_2 & \cdot & \cdot & x_n \end{bmatrix}}_{\text{design matrix}} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + \begin{bmatrix} e \end{bmatrix}$$



(That is a hemodynamic response example of the matrix which might help visualise).

- The first β (β_0) value and associated predictor (x) is a constant because the first x has a value of 1 for the whole time course (so the hemodynamic response is a straight line for the first predictor). This is so it can count as a baseline in which the fluctuations

$\boxed{Y = X\beta + e}$

of the other predictors can be modelled as increases or decreases relative to the baseline signal level (β_0 isn't ~~included~~ explicitly modelled in the equation because it's a constant but it is important to include it in the design matrix ~~A~~ - design matrix is where the predictors are modelled, so the Xs part of the matrix).

- The β values chosen for a model should minimise the error (reduce the amount of variance in Y which is left unexplained). So minimising the difference between the response predicted by the model and the actual response (Y) minimises the error of the model.
- To minimise the difference between the predicted data and actual data (Y) a statistical method that estimates ~~a~~ the optimal β values for the data set is used - called Ordinary Least Squares.
- So this GLM model predicts the components that made up the hemodynamic response in ~~one~~ a single voxel (or a single channel for us). The model can be repeated to get the components of the hemodynamic response for all the voxels (channels) of interest - the only thing that will change are the β values.

- Then we need to determine how well the model fits the data - because ~~the~~ even though our model might be the best predictor of the ~~ac~~ actual response (Y) it still might not significantly explain the measured response (Y).
- To determine how well the model fits our data (Y) the goodness-of-fit statistics need to be examined. This is measured from the multiple correlation coefficient (R value). The R value can be transformed to an F statistic which can give an error probability value ($p<.05$) of whether the ~~data is sign~~ actual data is significantly explained by one or more of the ~~β~~ model predictors (X).

- Statistical inferences - after the model parameters have been estimated and the model has been tested for its goodness-of-fit, specific hypotheses can be tested using t or F statistics. (e.g. to test whether two conditions significantly differ from each other)
- ~~These comparisons between conditions (or to baseline) using~~ These t or F values to make comparison between conditions (or to baseline) can be calculated using contrasts
- Contrasts are ~~weighted~~ combinations of beta values that correspond to the null hypotheses.
- E.g. if there are three conditions (so 3 predictors X that contribute/compose Y) and condition 1 has a ~~B~~ beta weight of 4.2, condition 2 has 0.6 and condition 3 has 3.5

$\boxed{y = x\beta + e}$

($\beta = [4.2, 0.6, 3.5]$) and we want to test whether condition 1 had more brain activity than condition 2 our contrast vector would be $[1, -1, 0]$
- we then multiply the contrast vector by the β values = $[1, -1, 0] \times [4.2, 0.6, 3.5] =$ $4.2 - 0.6 + 0 = 3.6$.
- The comparisons between conditions can take any form as long as the contrast vector sums to zero
- E.g., $[2, -1, -1]$ = ~~amore~~ testing whether there is more brain activity during condition 1 than conditions 2 and 3.
- ~~[illegible]~~ A t-statistic is obtained by dividing that value from the multiplication of the β values and the contrasts (in example 1 the value was 3.6) by the standard error of the contrast.
- To get an F value the contrasts can contain more than one row
- E.g. $\begin{bmatrix} 1, & -1, & 0 \\ 0, & 1, & -1 \end{bmatrix}$ = testing whether condition 1 is greater that condition 2 or if condition 2 is greater than condition 3.
- If we run comparisons for each voxel (channel) of interest there will be a high error rate - the multiple comparison problem. So this will need to be corrected for - eg. Bonferroni
- An unmodelled baseline is represented as 0 in the contrast vectors
- The statistical analysis will need a deeper understanding and wider understanding to be practically possible if using the GLM - this explanation just outlines the general idea about how it is done.
- Question - To compare conditions you take 1 β value for each condition and compare using contrasts. But that's just for 1 voxel/channel. What if we want to compare conditions for all of the voxels/channels in the ROI? Do we get (say by averaging) 1 β value for each condition over a ROI and compare that?

- PINTI'S GLM APPROACH - AIDE
- In the traditional GLM approach, the timings of the event onsets must be known in order to create box-car functions (and ~~for~~ event timings need to be known for averaging in order to compute the task and mean values)
- In lab based experiments this is ok because it is all controlled (stimuli onsets, trial order etc)
- But stimuli ~~[illegible]~~ presentation or event onsets in the real world aren't always controlled, which makes the traditional GLM approach problematic

- Even if using videos to record events, it might not be possible to tell ~~whether~~ when the hemodynam response happens ~~when~~ in response to the event - e.g. when they spotted ~~the~~ the stimulus.
- Therefore, Pinti has proposed a novel GLM-based method for the Automatic IDentificati of functional Events (AIDE).
- AIDE statistically detects functional events from fNIRS data (Y).
- Rather than taking the traditional approach of starting with a predetermined experimental design and investigating the effects of its events on hemodynamic activity, AIDE takes the opposite approach and starts with fNIRS hemodynamic response data and seeks to identify the ~~occurence~~ occurence of ~~events~~ event onsets from the measured data.
- The algorithm is based on the GLM model and identifies functional events by evaluating the best fit between different models of functional activity, assembled considering all the possible combinations of onset time and duration of the events and the measured fNIRS data.
- The algorithm has shown promising results but does not have the highest accuracy yet - maybe it would be the most useful tool in combination with other data forms (different measures additional to fNIRS)
- AIDE is the only form of analysis developed specifically for real-world neuroimaging.

REMOVAL OF PHYSIOLOGICAL NOISE MEASURED BY SHORT SEPARATION CHANNELS USING THE GLM
- If the GLM model does not contain all predictors (x) the measured response (Y) will be explained by the error values (e) instead of the model
- So it's important that the design matrix is constructed with everything ~~which~~. This may include sources of noise IF they have not been filtered out during pre-processing.
- If all of the predictors (x) have been properly included the error (e) should only reflect noise fluctuations.
- Gagnon et al (2011) created an algorithm that ~~no~~ uses ~~the~~ a GLM to recover the hemodynamic response ~~whilst~~ which simultaneously regresses out the physiological interference measured by short separation channels.
- The benefit of this is that getting information about the hemodynamic response is completed in one process compared to the two-step process of adaptive filtering (the adaptive filter and then averaging).
- Gangon also found it to be the most accurate method of hemodynamic response recovery and it is used ~~by~~ by researchers now.

- Some papers do use this method and then go on to use block areaging to compare condition

OPTODE PLACEMENT

- Unlike functional MRI, fNIRS does not provide structural information about the brain
- This means that identifying active brain areas, ~~or~~ and ~~compared~~ comparing or reproducing results between studies can be a problem.
- So instead of taking MRI scans of each participant and ~~exact~~ precisely identifying brain activation from the fNIRS data, a standardised estimate of structural activation can be used in fNIRS
- The Montreal Neurological Institute (MNI) is a brain atlas - it is the average of several healthy brains to make a 'standard' brain
- In EEG, electrodes are placed on the scalp according to the international 10-20 system of electrode placement
- This system ~~utilises~~ utilises external landmarks on the scalp to determine the location of underlying cerebral structures (brain areas)
- In fNIRS, the ~~10-20~~ international 10-20 system can be used - placing the optodes in accordance with the system means that results can be compared between studies AND the activated brain areas can be estimated
- The brain areas activated can be estimated because ~~they the optod~~ the system corresponds to the MNI atlas - see Okamoto et al. (2004) and Cutini et al. (2011) for brain areas (and Brodmann's areas) ~~for brain areas~~ that correspond ~~to~~ between the location ~~of~~ of the international 10-20 system and the MNI ~~at~~ atlas.
- Each placement area in the system has a letter to identify the lobe or general area of the brain - prefrontal = Fp, frontal = F, temporal = T, parietal = P, occipital = O, and central = C. ~~Even numbers~~
- When positioning the cap, we want CZ to be at the location that is halfway between the nasion and inion and halfway between the preauricular and preauricular (part of the ears) - ~~so~~ this shouldn't be a ~~problem~~ problem if the cap size is correct. With the headband fNIRS cap manually measure the placement area positions in relation to the CZ position and place the headband accordingly
- So just make sure that the optodes ~~cover~~ are in the positions of the international 10-20 system areas and then you can compare results with other studies and make assumptions about which brain areas were activated.