An Interdisciplinary Exploration of Information Overload

Mohamed Amine Belabbes

Information Engagement Research Group
Department of Computer & Information Sciences
University of Strathclyde, Glasgow

A thesis submitted for the degree of

Doctor of Philosophy

October 2022

Declaration of Authenticity and Author's Rights

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Signed: Mohamed Amine Belabbes

Date: 05/10/2022

Acknowledgements

I would like to thank two individuals without whom I would not have made it to the end: Dr. Diane Pennington and Dr. Gemma Learmonth. To me, they are the two greatest supervisors ever and who made the last year the best year of this long journey. Diane, you took me up during one of the most difficult times in my life and you have shown me great respect, appreciation, and love. I believe these three are the best combination that people can receive. Thank you for your guidance as I was putting together my thesis, thank you for teaching me the fundamentals, thank you for putting me in a good context and welcoming me to your research group, and thank you for your faith in me. Now, Gemma, the superhuman, you engaged with my work and ensured I was progressing throughout my PhD through more than half of your maternity leave...It meant a lot to me. Your dedication is one thing I will take away with me, something I will always look up to. Diane, Gemma, I have said this, and I will keep saying it again and again: From the bottom of my heart, thank you. If there is ever a hall of fame of supervisors, well you should be up there. I would like to thank Prof. Frank Pollick, Dr. Yashar Moshfeghi, and Prof. Ian Ruthven for their academic advice during this journey. I would also like to thank my friend Tina for taking the time to read again and again my thesis.

Without my parents, none of this would be possible. I would like to thank my dad and my mom for all the sacrifices they have made for me and my sisters. Throughout all the challenges, they have been patient, and supported me. I look up to you two, you are the best. I love you!

So many people kept me company throughout this journey. They say “it took a village”, well, it might as well have been the case for me. So many people have shown care, support, and love throughout this journey – people who did not have to, yet they came into my life and brought something positive. I hope this work can make you proud of me. Thank you.

To Dad, Mom, Meriem, Sarah, Nabil & Charles

To everyone who believed in me when I didn't

To everyone who has shown kindness

To everyone who has shown love

Contents

List of Figures

List of Tables

Abstract

While the propagation of information can lead to innovation and reduce inequality, the fast propagation of information, and the lack of gatekeepers which can support individuals online has led to the proliferation of the problem of information overload. The problem of information overload has always been with us, however, with the development of the world wide web and of new applications serving information, new dimensions to information overload appeared. Attempts have been made to tackle it through various approaches, the most notable being recommender systems. Yet, the lack of a clear definition makes it difficult to scope the problem and clearly identify which issue is being tackled and whether it is tackled efficiently. In the academic context, recommender systems are usually evaluated on past datasets.

Hence, using a method from nursing science called concept analysis, we provide a clear definition of information overload. The concept analysis offers a clear outline composed of causes, manifestations, and consequences of information overload. This breakdown enables better scoping of the problem, and the operationalisation of the concept with the aim to conduct more empirical studies.

Our work further provided outcome measures which can lead to a scalable way for future systems to detect information overload. These outcome measures were identified during a user study based on the social network Twitter. They were derived from the interaction of our participants with our Twitter clone. By leveraging machine learning methods, we developed classifiers to detect whether individuals were under IO, these classifiers can reach up to 80% accuracy. We found that outcome measures related to mouse tracking solely could lead to detect when individuals were under IO.

Our work also showed that while individuals might feel information overload through its emotional and cognitive manifestations, it is not a phenomenon which affects their

behaviour instantly. Rather there is a time gap before its onset and its influence on the behaviour of the individuals. We established that individuals can identify that they are in fact subject to information overload; however, there is a delay before it affects them.

Complementing our attempt at tackling IO through developing smarter systems, this work investigated how information overload can be tackled from the side of the users of a system. To do so, we designed a search system (a clone of Google) and leveraged advances in cognitive neuroscience and in the research area of neuromodulation to stimulate a brain region which was previously shown to be involved in the online search process. The brain region stimulated was the left dorsolateral prefrontal cortex, part of the prefrontal cortex which is associated with executive functions. Our work highlighted the important considerations needed to be given to executive functions when studying the concept of information overload. Finally, for both of our Twitter-based study and the search-based study, we open-source both systems to offer the research community the opportunity to develop further studies using similar experimental paradigms to those that are used in this thesis.

Chapter 1

Introduction

With the exciting growth of the internet, information has become easily accessible to individuals all over the world. It has allowed us to enter a new era, which is commonly described as the 'age of information' [3] or the fourth revolution [4]. This fast development of information was enabled through important developments in the field of information retrieval (IR) [5, 6], and such ease of access to information has enabled tremendous progress and development of new opportunities for individuals [7].

A side-effect of this revolution is information overload (IO). IO is a problem which has been present throughout history; its first mention can be traced back to Biblical writers [8]. However, it has seen a change in its nature, particularly with the development of the internet and social network sites (SNS). Today we find it more present than ever with new causes as well as new consequences [9]. Yet, in fields such as IR, we find literature aimed at tackling IO without providing a definition beforehand or clearly outlining which deficiencies are being investigated, but also without taking into account the individual differences of the users. There is an assumption that IO is a broadly discussed and understood topic and, while in its primitive form this might be the case (i.e quantity of information), there are more dimensions to IO, such as the information environment [10].

This thesis looks at the problem from a computer & information science (CIS) perspective and aimed to clarify the concept of IO by providing a robust definition and investigating how studies on the subject can be improved.

1.1 Motivation

In the field of information science, a common problem when addressing the fundamental concepts of the field exists which is the lack of consensus on the meaning of said concepts and their scope. [11, 12].

This can lead to various challenges when performing empirical work. Jarvelin and Vakkari [13] acknowledged that there has been empirical work aimed at bringing conceptual clarity to the field's concepts, however, this work can not be a substitute to an analysis of the basic concepts composing the field. Such shortcomings lead to a scattered effort when studying a concept.

When Fleming-May [14] discussed the problem of conceptual clarity in information science, they mostly focused on the popular concepts associated with information seeking. One concept that is missing and often poorly discussed in depth in computer & information science (CIS) is the concept of information overload (IO). In the field of CIS and particularly information retrieval (IR), IO is often mentioned with the assumption that there is a consensus on what is being discussed. While it is safe to assume that in its most popular form, we are discussing the problem of the large quantity of information, there have been new dimensions which have appeared due to the development of new information systems and due to the complexity of some.

While content is now constantly produced, the challenge of filtering through it is no longer only associated with the goal of acquiring knowledge but in some cases with passively browsing content on social network sites (SNS). This approach to sifting through information is causing issues today as policy-makers become concerned, for example, with the impact of the consumption of information that can have on children, such as addiction [15].

When browsing information online, we rely on what is denoted as executive functions. These executive functions allow us to make decisions, store information, and perform associations. As IO became a fundamental component of the "fourth revolution" [4], at the individual level these executive functions were tested.

In this thesis, we take a first step towards understanding the causes of the aforementioned problems. This step involved the disambiguation of the problem and exploration of solutions.

1.2 Thesis statement

Information overload (IO) has been tackled in the past in computer & information science (CIS), particularly with the advances made in artificial intelligence, and in machine learning algorithms used for recommendation. However, it is only mentioned as a concept with the belief that everyone knows what is being addressed. Such issue in defining fundamental concepts of a field is common in information science [14]. A clear definition of the concept would allow to not only better tackle it, but also contribute to the theoretical underpinnings of the broader field of research, in this case CIS. Hence, we investigated the use of the terms IO and cognitive overload (CO), in order to bring more clarity to the meaning of these terms and provide a definition. While IO is a broad concept, being able to highlight which aspects are to be addressed can help to better understand **how** they could be addressed. Understanding the components that are part of IO can also help design studies in a controlled environment. This led us to design two studies, in each study we addressed different triggers to IO. One study focused on the information characteristics [9] and approached the problem of IO from a system-side view. The second study focused on the cognitive abilities of the individuals [9], hence, this led us to look at work from cognitive neuroscience.

In this thesis, we aimed to show the overreaching presence of IO through its various dimensions. Through our review of IO, we highlighted the negative manifestations and consequences which result from IO, and impact both the individuals and their environment. The principal question we asked is ‘What is information overload?’, to understand the problem and define it, we took this first step which led to our secondary question ‘what are the possible approaches to tackle it?’.

To answer these questions we defined eight research questions (section 1.3) which were resolved through a combination of theoretical and empirical work. We focused on resolving the first question by defining IO using a reproducible approach called concept analysis [16], subsequently we used those findings to test our definition of IO. This led us to explore how IO impacts users' behaviour and to go further by tackling the problem at a visceral level by investigating the use of brain stimulation to support users of search systems. Our empirical work used two systems, one clone of Twitter and one clone of Google. Both systems were used for our studies, we open-sourced the code (subsection 3.3.7 and subsection 4.2.7) to encourage future work to be reproduced in a similar environment, and make future findings more robust.

1.3 Research questions

The aim of this work is to offer a conceptual clarification of information overload and explore the potential approaches to be taken to tackle it.

Our first contribution resides in our definition of information overload, and showing that 'information overload' and 'cognitive overload' are the same concepts in computer & information science. Based on our definition of information overload (IO), our subsequent contribution is in characterising the behaviour of individuals under IO and under not IO. This led us to an additional contribution based on the changes in behaviour: Leveraging the behavioural changes of individuals under IO and not IO, we were able to highlight the evolutionary nature of IO across time, and identify features for the design of scalable IO classification models. Our final contribution consists of the exploration of the use of neuromodulation in computer & information science (CIS), we attempted to reduce individuals' dwell time (time spent on web pages excluding the time spent on the search engine home and results' web pages). We found that a 2 mA stimulation using transcranial direct current stimulation (tDCS) for 15 minutes on the left dorsolateral prefrontal cortex did not lead to a reduction of the dwell time. However, we offer guidance on how to design future studies and avoid confounding factors such as tDCS side effects, and our experience can benefit both the CIS community and the cognitive neuroscience community.

The main research questions of this thesis are:

RQ 1: What is information overload?

RQ 2: What, if any is the difference between information overload and cognitive overload?

RQ 3: What is the behaviour of individuals under information overload when interacting with an information retrieval system?

RQ 4: What is the behaviour of individuals which are not under information overload when interacting with an information retrieval system?

RQ 5: How is information overload perceived over time?

RQ 6: What are the important features to develop a session-based information overload classifier?

RQ 7: Can we reduce the dwell time of individuals on an online search task using neuromodulation?

RQ 8: What is the impact of the left dorsolateral prefrontal cortex stimulation on the interaction of individuals with an information retrieval system?

1.4 Thesis outline

This thesis is constituted of five chapters, the chapters are organised as follows:

Chapter 1 - Introduction

The first chapter presented the motivation of the research, the knowledge gap that this research has aimed to fill, and presented our thesis statement.

Chapter 2 - Information Overload: A concept analysis

In this chapter, we contextualised the problem of information overload (IO) within computer & information science (CIS) and showed how it became not only an information science problem but also a computer science problem with the development of the field of information retrieval (IR). Through this contextualisation of IO and by highlighting its development, we demonstrated the lack of a formal definition of IO. Hence, we used a proven method to achieve conceptual clarity namely concept analysis to define IO [14], and provided a framework of IO to enable more empirical work.

Chapter 3 - Recreating information overload in the lab: A Twitter study

In this chapter, we designed a user study which is based on social network sites (SNS), and focused on recreating IO in a controlled environment. We used the findings from our concept analysis to verify whether we can successfully create IO in a controlled experimental setting based on our framework while exhibiting relevant manifestations which can be used to develop a subjective measure of IO. Finally, by replicating approaches to study behavioural changes, in this case mouse data [17], we extracted implicit behavioural changes of our participants across the IO and not IO condition. Based on our log data, we derived outcome measures and evaluated which ones are the most affected when users of a social network are under IO. These outcome measures are used for feature engineering, and the development of a model which can classify whether users are under IO or not.

Chapter 4 - Neuromodulation: A novel approach to tackling information overload

In this chapter, we aimed to tackle IO from the users' perspective by modulating the brain regions that are responsible for executive control. We designed a user study in the context of search engines, more precisely in the context of exploratory search [18] and developed a clone of Google to recreate a high-fidelity search experience. In this study, we used a neuromodulation technique called transcranial direct current stimulation to reduce IO during an information retrieval task. We looked at the dwell time on pages that were consulted to detect whether users of the search engine improved based on the behavioural data after anodal (excitatory) stimulation of the prefrontal cortex.

Chapter 5 - Discussion & conclusions

In this chapter, we discussed the contributions of this thesis, presented areas for future work, and highlighted the implications of our findings for the research community, engineers, as well as policy-makers.

The findings in each chapter (chapter 2, chapter 3, and chapter 4) are intended to be self-contained, therefore some information may be repeated across chapters.

1.5 Publications

In this section, we list the work which was accepted to journals and peer-reviewed conferences. Each publication refers to parts of the work done as part of this thesis.

- Belabbes, M.A., Ruthven, I., Moshfeghi, Y. and Rasmussen Pennington, D. (2022), "Information overload: a concept analysis", Journal of Documentation, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JD-06-2021-0118

Chapter 2

Information Overload: A concept analysis

In this chapter, we started by providing some context around information overload (IO) in computer and information science (CIS) and highlighted the attempts made by the community to address the problem. Subsequently, we reviewed how the problem of IO is perceived in CIS and how it became a problem of the information retrieval (IR) community but also the CIS community in general. In the background, we also presented past work in CIS which looked at using the method of concept analysis to clarify concepts of the field. In the methodology, we introduced the approach we used to perform our concept analysis and how we built our dataset. Finally, based on our concept analysis we provided a definition of IO and find that in the CIS community, IO and cognitive overload (CO) are one and the same.

2.1 Introduction

Information overload (IO) has become an omnipresent and persistent problem due to the ease of access to information enabled by the internet. While in the past, IO might have been perceived as a problem resulting from the large quantity of information [19], nowadays IO encompasses more than a problem resulting solely from the quantity of information available.

For such a consequent problem, one would assume there would be a definition that researchers could agree on as they study the problem; however, there is no consensus on one specific definition. One popular definition in information science is by Bawden et al. [20]: They define IO as a consequence of a large quantity of useful and relevant information that it becomes an obstacle. In parallel to the lack of consensus on a definition of IO, we find that another term is often used interchangeably, 'cognitive overload': This term comes from cognitive psychology and is linked to Cognitive Load Theory (CLT) [21]. In CLT, Sweller [21] defines cognitive overload (CO) as the result of overloading working memory (a type of short-term memory, see chapter 4) with information.

In recent years, in the field of computer and information science (CIS), researchers have paid a closer attention to the problem of IO: Within the ACM digital library in 2003, the keywords 'information overload' and 'cognitive overload' were mentioned in 53 publications. While in 2020, 'information overload' and 'cognitive overload' were mentioned in 254 publications.

As we left the industrial age, we entered an information-based economy [22] which led to new spaces where IO can develop. Various elements can explain this linear growth of researchers' interest in IO, among them is the development of social network sites (SNS): According to Koroleva et al. [23]'s conceptual model, IO in SNS is positively correlated with the increasing number of connections in a user's network. In addition, they noted an increase in users discontinuing their use of a service as a means of coping [23, 24, 25]. Due to their nature, SNS have all the attributes for the development of IO amongst its users: Diversity of content and modality used to share information (text, images, videos), irrelevant information, pushed advertisements, and a large number of (sometimes conflicting) sources [25, 26].

The field of information retrieval (IR) was born from the need to index and retrieve documents due to the large quantity of information that was becoming available [19], it drove the development of various algorithms such as term frequency–inverse document frequency (tf-idf, more about the nature of the field of IR in chapter 4). Tf-idf allowed to speed up the indexing process by attributing weights to the words present in a document and using the most important words as keywords for a given document [27]. However, prior to tf-idf, early work had only achieved poor recall (fraction of retrieved relevant documents among all the relevant documents available) and poor precision (fraction of relevant documents among the retrieved ones). This inefficiency in retrieving relevant documents due to the shortcomings of the algorithms developed was an early cause of IO in the field [28]. As more sophisticated algorithms were developed and as the ease of access to various sources increased, IO's challenges became more diverse: Poor information literacy, poorly defined information need, poor language proficiency, increase in duplicate information, increase in redundant information, increase in the quantity of information, individual limitations related to their cognitive abilities (limited working memory), and external factors related to time given to complete a task and the task complexity [29, 30, 31, 32].

These challenges led to the development of solutions that can allow for a reduction of the information to be explored by individuals to satisfy their information need. These solutions have consisted of new filtering algorithms, and with the developments in the field of artificial intelligence we have alternatives to the traditional search process that appeared: Recommender systems which can offer suggestions based on individuals' past search trail, machine learning based summarisation method [33, 34, 35, 36]. In addition, various research has been looking at the development of emotion-based feedback systems which can identify the emotions elicited by information to evaluate its relevance [37, 38, 39].

Prior to the boom in artificial intelligence, Montebello [28] in 1998 argued in his paper that IO resulting from searching on the internet is solely an IR problem and they presented a functional system which leveraged artificial intelligence and integrated elements that are now commonly incorporated in IR systems such as user profiles. Until recently, IO in IR was not as much approached from a user oriented view. The approach of IR to study IO contrasts with other fields such as organisation science, marketing, and other disciplines where more studies were performed on IO with more variety in dimensions explored and more attempts at defining it were made [40].

However, a missing framework of IO with its definition is lacking computer & information science (CIS), it would allow the better scoping of the developed solutions. A few dimensions that have become pressing matters to solve are: The growth of misinformation accompanied with the spread of fake news on social media platforms [41], the type of information shared can lead to the exacerbation of poor mental health amongst the broader popular [42] which can result in avoidance behaviour and poor performance [43]. Finally, social media platforms became sources of information for organisations such as emergency services that rely on them in crisis (e.g., natural disasters, attacks, etc.): These emergency service providers must therefore rely on their attention to filter out unreliable content and speculation [35].

Hence, a disambiguation of the concept of IO can allow for the concept to be better studied by academia but also for engineers to develop better software, and for policy-makers to identify its dimensions and encourage industrial partners to take account of it when developing technology. A clear operationalisation will lead to greater academic rigour when investigating the problem with the intention of developing solutions, and most importantly being able to evaluate these solutions before their deployment and use by industry. This lack of conceptual clarity is common in information science, and it has been a cause of concern with the field's foundational concepts [11, 14]. To achieve the disambiguation of IO, we borrow from nursing science and perform a concept analysis [16] based on Rodgers' approach [2]. Rodgers' approach offers a clear, and well-documented method to clarify and define concepts [44].

In the literature, we find various terms which have been used interchangeably with information overload [45], a common one is cognitive overload. As highlighted, 'cognitive overload' is a term that comes from cognitive science, it is commonly used when referring to Cognitive Load Theory [21]. Yet, in computer & information science, we find numerous papers using the term of cognitive overload in conjunction with information overload [46, 47]. This leads to question whether information overload and cognitive overload are the same concept. Therefore, a disambiguation of information overload requires to also look at cognitive overload: For our concept analysis, we build our dataset with papers discussing information overload and cognitive overload. Considering both terms allows to see if they share overlapping themes and allows to cover all the causes and consequences of IO. In this chapter, we answer the first two research questions of this thesis:

- **RQ 1**: What is information overload?
- **RQ 2**: What, if any is the difference between information overload and cognitive overload?

The following chapter is organised as follows: In the next section (section 2.2), we highlight why this concept analysis is necessary by going over past research and over the origins of the term 'information overload'. In addition, we review the method of concept analysis to show how past work has efficiently been used for the disambiguation of various concepts. Subsequently, in our methodology section (section 2.3) we present our approach to concept analysis as well as the procedure we followed to build our dataset. Based on the analysis of the dataset, we present our findings (section 2.4) which consist of a definition of and a framework to study IO. Finally, we conclude by discussing our findings and their implications (section 2.5).

2.2 Background

In this section, we start by presenting how information overload (IO) developed throughout history and use the context as a bridge to how it became a problem for the field of IR and more in general the field of computer & information science (CIS). We also extend our background with a brief review of recent work in CIS that used concept analysis to study concepts.

2.2.1 The development of information overload

The first allusions to information overload date back to Biblical writers and Socrates [48]. The presence in history of the concept indicates that the problem is not a modern problem. The French philosopher Diderot alluded to the problem when they predicted that with time, the number of books will keep growing that eventually finding information will be as challenging as studying the universe directly [49]. Through their observation of the spread of information through the press and the telegraph in the 19^{th} century, Beard [50] was arguably the first to describe the consequences of IO as fatigue and anxiety; In their book titled 'American nervousness, its causes and consequences: a supplement to nervous exhaustion (neurasthenia)', Beard identified the medical condition of neurasthenia as a consequence of the propagation of information and the overload it creates. Neurasthenia is associated with the modernisation of society and while by today's standard it is an ill-defined medical condition, it is used to describe a drained nervous system [51].

The first mention of a need to develop a system to tackle the growing amount of information dates to the end of the Second World War: Bush [19] suggested the development of a machine called the 'memex' to tackle IO, such machine would allow people to organise books and documents. The memex could thus be considered as the first description of a potential information retrieval system. In this context, the dimension of IO that was of concern is the quantity of information available due to the accelerated scientific developments that were happening at the time.

However, the term of IO is more recent: Levy [52] notes it was first used by Meier [53] in the field of urban studies. While Sasaki et al. [24] credits Gross [54] with the initial application of the term 'information overload' in the field of organisation science. While there is no consensus on who used it first, the term was only later on popularised by Toffler [55] in his book *Future Shock*.

In organisation science, Gross [54] coined the term of IO and described the concept as a result of individuals being overloaded with information leading to poor decision-making. Virkus et al. [56] considered the first study to focus on IO as Miller [57]'s, it is also often referenced when attempting to define IO. In their study, Miller [57] looked at the limits of human working memory by investigating the amount of numbers individuals can remember before their recall suffers. Another definition of IO was by Schroder et al. [58], they defined it through its consequences and associated IO with cognitive overload (CO): As the quantity of information being processed by an individual increases, at a certain time utility decreases ; it is often depicted by an inverted U-curve.

In their work, Bawden et al. [20] defined IO similarly to how Schroder et al. [58] did, IO is the result of an accumulation of information that it becomes a hindrance rather than a help. Later on, Bawden and Robinson [32] highlighted that while the nature of IO does not change, there have been new causes: novelty, diversity, complexity, the nature of how information is faced (pushed vs pulled), as well as individual differences. The latter cause of IO is presented by Savolainen [59], they described IO as a phenomenon that older people are more prone to suffer from rather than it being some cognitive difference across old and young. Furthermore, Savolainen [59] presented it as a result of information literacy of younger people who are more skilled at using technology to access information. A related additional cause of IO is the lack of efficient filtering mechanisms, while in the past there were gatekeepers (e.g librarians) which could help in filtering and ensuring the quality of information, today anyone is free to directly satisfy their information need on their own [26]. Similarly, this free access to information is also a result to the development of social network sites (SNS), as the number of users increases, there is more to crawl through and this passive consumption of information can lead to IO [60]. In parallel to the growth of sources of information on SNS, more and more sources of information have become available on the internet: Such overall increase has led to unfiltered information, information without quality control can be redundant, hence putting an additional cognitive burden on the individuals accessing the information [25, 46, 60, 61, 62]

New causes of IO were accompanied by new consequences, one of the consequences mentioned by Bawden and Robinson [32] is the impact on mental health, as well as attention deficit trait [63] and "cognitive overload" [64]. [63] described attention deficit trait as individuals reaching the limits of their executive functions (see chapter 4 for definition) and hence finding difficulties staying organized, setting priorities, and managing their time. Kirsh [64] characterised IO as information which is too large in quantity, hence IO is one cause of CO, with additional causes being: overload of information demand, increase need for multi-tasking, the environment causing interruptions and poor workplace infrastructure. As we progress in this work, we will find that in CIS, these causes and consequences all constitute IO. In our review of the new consequences of

IO, we found that individuals will adopt a behaviour of avoidance in regards to media with redundant information [65] which would constitute of either a filtering strategy or completely withdrawing from a task [59]. In some cases, the choice of media from which to receive information cannot be refined due to time constraint such as in disaster situations where first responders have to extract valid information to understand the impact of the disaster [29, 35, 66]: At a visceral level, this would require an increase in attention and an increase in the individual's working memory processing demand. We found a similar demand in collaborative environments, where we have various actors contributing to a knowledge base and it requires an active filtering to access relevant information [67]. In the case of e-commerce systems, extensive product descriptions can lead individuals to poor rational in buying choices associated with poor decision-making [68]. Finally, news reporting suffers from the increases of sources on SNS and hence, this results in difficulty in filtering and the propagation of fake news [35, 69]. This ease of news propagation by anyone on SNS and the increase of sources results in high exposures by the users to news items, which then results in stress [70].

The diversity of the causes and consequences of IO leads to a diversity of terms used to describe IO: Information anxiety, infobesity, communication overload, and cognitive overload [8, 30, 71, 72]. While some researchers have used information overload (IO) as an umbrella term for one aspect of cognitive overload [64], others such as Marques and Batista [72] consider it as an attribute of communication overload due to the increase of communications means (the web and the elements composing it being one) between individuals and groups [56, 72].

The diversity of terms used exacerbates the ambiguity that IO suffers from and highlights the lack of consensus on what the concept is and the dimensions that constitute it [32, 73].

2.2.2 Concept analysis

Conceptual clarity is necessary to establish the theoretical foundations of a field of study. However, as highlighted by Fleming-May [14], the field of library and information science has struggled with this. The lack of conceptual clarity does not only affect IO, but in information science as stressed by Vakkari [12], it also affects more basic concepts related to the information seeking behaviour. Vakkari [12] called attention to authors of publications at the time who would use concepts such as information need, knowledge, and information seeking without a specific definition: They added that these concepts are used and “take their meaning as given”. A similar phenomenon to what we have observed in the case of IO, which prompted this work. Hence, we decided to use a concept analysis as our method to achieve the operationalisation of the concept.

Concept analysis is a method that is most often applied in the field of nursing science to achieve the disambiguation of concepts, and the consolidation of a concept across the field [14, 16]. In reviewing concept analysis as a candidate method for clarifying notions in library and information science, Fleming-May [14] found that the reproducible frameworks that conceptual analysis benefits from could help the field in improving their knowledge when studying concepts, establishing fundamental definitions of concepts, hence reaching consensus in what is being studied and allowing the community to perform more empirical work on the concept.

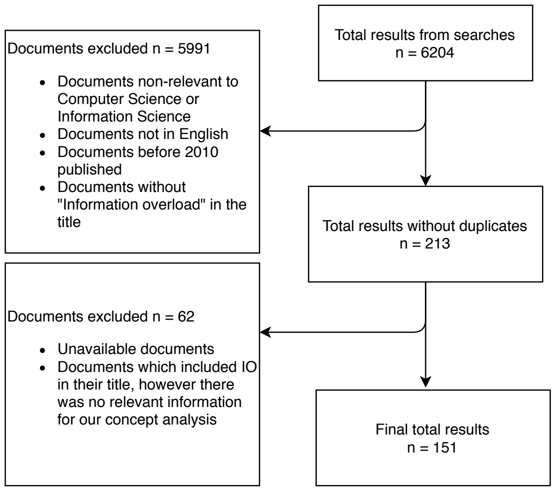
One use of concept analysis in information science was by Savolainen [74] who were able to shift from the vision of strategies as a plan of action, to a disambiguation of the concept of strategies and breaking it down to two types: deliberate and emergent strategies. Deliberate strategies are intentionally adopted, while emergent strategies are the result of an adaptive mechanism of the individual facing information as a result of the absence of deliberate strategies or regardless of their presence. Additional work in library and information science used concept analysis: Savolainen [75] employed it to identify how the concept of information need differs based on its context, work has looked at concept analysis for the concept of health information seeking [76, 77], and library use [78]. Savolainen [79] performed a concept analysis on the cognitive barriers of information seeking and highlighted how information overload is one of them: They present two outcomes of the overload, outcome overload and textual overload. The outcome overload is described as the inability to go through the list of results available and evaluate their relevance, while textual overload represents the inability to read the identified relevant corpus. While in their paper, Savolainen [79] mentioned the deterioration of attention and working memory for elders, they only referenced it as an obstacle to articulate an information need. In this thesis, following the clarification of the concept of IO, we went further and identified what it means at a neurological level to be overloaded by recognising neural correlates involved in the online search process, and stimulating one specific brain region to try and tackle IO from the users' side (chapter 4).

2.3 Methodology

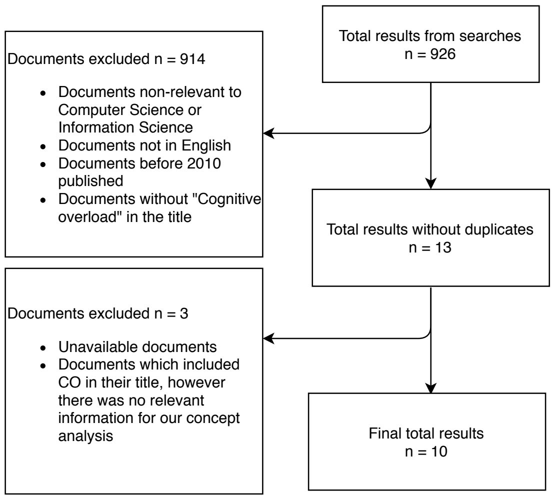
2.3.1 Inclusion and exclusion criteria

Since our concept analysis covered the fields of computer science and information science, we chose three databases: ACM Digital Library, LISA, and SCOPUS. The ACM digital library is a database of literature of the fields of computer and information technology, LISA is an important database for information science, and SCOPUS a more general database which offers articles from across various fields. In each of the databases, we applied a filter by field with "Computer Science" and "Information science" selected.

During our initial phase, we retrieved 6204 documents for IO and 926 for CO based on the following inclusion criteria: (1) 'information overload' OR 'cognitive overload' as our query, (2) literature published between 2010 to September 2020, (3) written in English, (4) published in CIS. Due to the initial large sample of documents returned, we had to limit the results to only the documents that contained our search terms in the title. Constraining the results to the ones with our search terms in the title led 213 documents for IO and 13 for CO without duplicates. A final filtering led to removing unavailable documents as well as documents which had the search terms in the title but did not have relevant information (Figure 2.1).



(a) Information overload literature



(b) Cognitive overload literature

Figure 2.1: Survey flow

2.3.2 Dataset

After applying our inclusion and exclusion criteria, our dataset consisted of 151 documents for information overload (IO) and 10 document for cognitive overload (CO). Based on an initial review of the literature, we found that the most common surrogate of IO is ‘cognitive overload’, hence our choice for the term. All of the papers reviewed were from the fields of computer & information science. During the core analysis phase, we identified ‘information anxiety’ and ‘infobesity’ as two additional surrogate terms due to their recurrence within our dataset. Our inclusion and exclusion criteria applied for both additional surrogates, we found 7 papers for ‘information anxiety’ and 1 paper for ‘infobesity’: We found that infobesity shared the same themes as IO [80], while ‘information anxiety’ had non-overlapping themes with IO therefore we considered it as a related term rather than a surrogate of IO.

2.3.3 Rodgers’ approach

The choice of method to clarify a concept has to fit the concept being studied, the overreaching goal in such cases is to be able to scope the concept and highlight its dimensions: its causes and consequences as per the literature. So far we have highlighted some of the forms in which IO appears in CIS, and how various terms were used to describe it: choice overload, cognitive overload, infobesity, information anxiety...etc. This plurality of terms can lead to a misunderstanding amongst researchers about whether it is the same concept being studied or not. In order to achieve a consensus within the field and contribute to the disambiguation of IO, we aimed to use an approach which is reproducible and would allow us to define the concept. Hence, for IO we chose to perform a concept analysis. A concept analysis is one approach to study concepts by reading the literature and examining their attributes, thus contributing to the development of new theories. These theories could then be further tested through empirical work [16]. The method we use to perform our concept analysis is Rodgers’ approach [2]. Rodgers [2] describes it as an inductive approach, rather than offering definite conclusions, it offers opportunities for the researcher to further develop a concept: It consists of an initial phase, a core analysis phase, and a further development phase. Rodgers’ approach is systematic, and all these three steps can be repeated and done in parallel.

Surrogate	What are the terms used interchangeably with the concept? (if any)
Related terms	What are the terms which share overlapping themes with the concept? (if any)
Antecedents	What are the events that lead to the concept?
Attributes	What are the characteristics of the concepts?
Consequences	What is the result of the concept?

Table 2.1: Items to extract Rodgers [2] approach to CA

In the initial phase: The researcher identifies the concept, determines the research areas, the various terms associated with the concept, the inclusion and exclusion criteria, and the databases are chosen. By the end of this phase, we have determined the collection of documents which constitute our dataset.

In the core analysis phase, as the researcher reads their collection of documents, they have to identify pattern and define the main themes that are present. To guide the researcher in the analysis, there are five elements to extract from each document: (1) Surrogates, (2) related terms, (3) antecedents, (4) attributes, and (5) consequences. Each of these elements is associated with questions, see Table 2.1 [44]. During this phase, if there are additional relevant surrogates then the same process that was followed is repeated, the dataset is expanded using the surrogates identified as the keywords.

The further development phase allows the researcher to use the identified patterns and themes to offer further questions and hypothesis to be further answered by the community. The themes allow establishing connections across the concept, this leads to defining the concept. Once this final stage is completed, the definition and findings provided can then be used by the rest of the community to further test the findings but also perform further studies to confirm them.

2.3.4 Procedure

During the first stage, all the documents of our dataset were read in their entirety by the principal researcher for the first time, and a second time where the questions from

Rodgers' approach were answered (Table 2.1). We extracted the surrogates, related terms, antecedents, attributes, and consequences. These were all put together in an Excel spreadsheet, each row represented a document from the dataset.

During the second stage, the principal researcher went across the data extracted and identified initial themes: Brain ability and cognition, information need, mental health, intrinsic and extrinsic information characteristics, and internal and external consequences. Subsequently, an additional review of our data led to the final step of refining these themes which resulted in the main themes presented in our framework with their groups: triggers, manifestations, and consequences (see Figure 2.3). We breakdown our framework and findings in our results section (section 2.4).

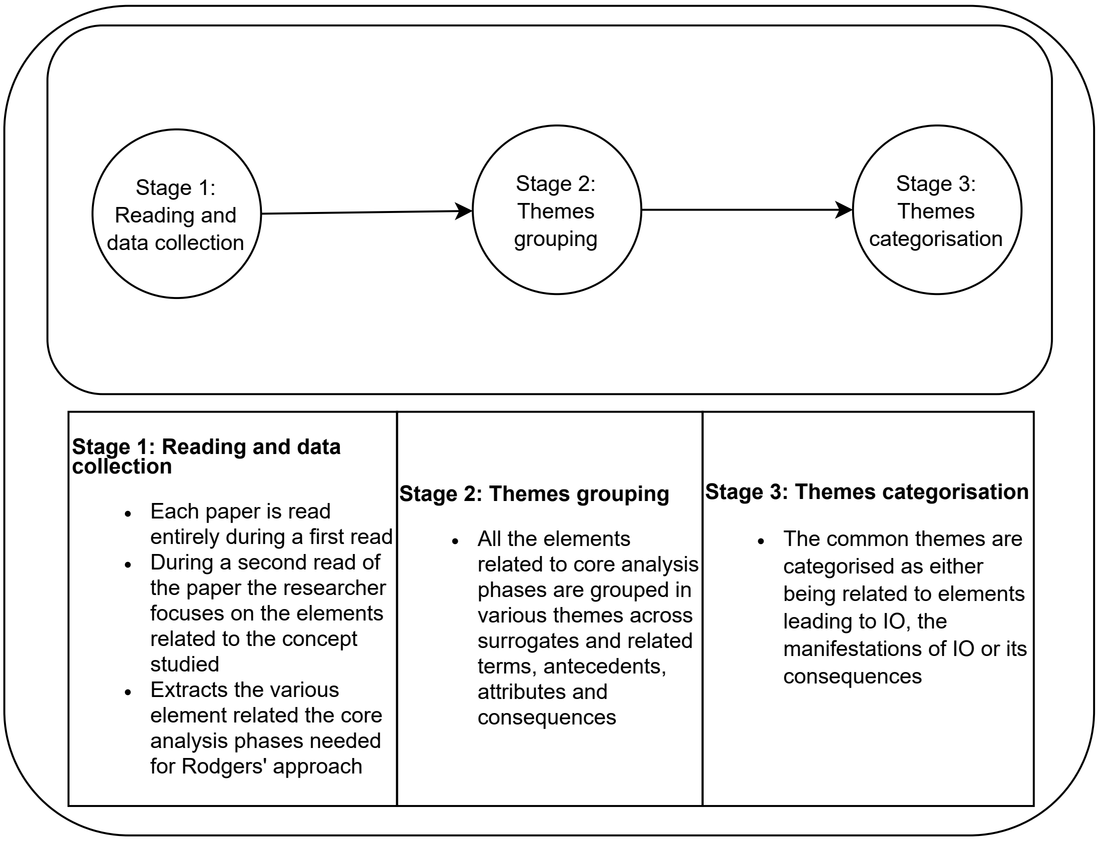


Figure 2.2: Concept analysis process

2.4 Results

The results section presents our findings from applying Rodgers' approach to our concept analysis on the concept of information overload (IO), we were able to establish the different elements that constitute IO, how they relate to one another, and how they shape IO.

The first finding as a result of using Rodgers' approach was during the core analysis phase, all the themes that were mentioned in the cognitive overload (CO) dataset were the same as IO, and papers have used them interchangeably. Hence our first conclusion is that IO and CO are one and the same concept in the field of CIS. This finding led us to combining IO and CO during the further development phase and establishing the main themes which constitute the concept of IO (Figure 2.3); the themes were grouped under three categories: (1) Triggers which constitute the causes that lead to IO, (2) manifestations are the elements that happen within the individual when experiencing IO, we found emotional and cognitive manifestations, (3) consequences which are the elements that happened as a result of IO and can be both internal (i.e linked to the individual's internal state) or external (i.e how the individual experiencing IO impacts the world).

In this section, all the work referenced was part of our dataset.

Triggers

As part of our analysis into the causes of IO, we identified five main themes that we group under 'triggers'. These triggers are associated with the way information is presented, the nature of the information presented, the individual's cognitive ability, the nature of the individual's information need, and finally the context (i.e environment) in which the individuals find themselves when collecting information.

Intrinsic and extrinsic information characteristics.

Two important factors which can determine whether an individual will be subject to IO are the intrinsic and extrinsic characteristics of information: The intrinsic characteristics of information are related to the nature of the information presented (i.e

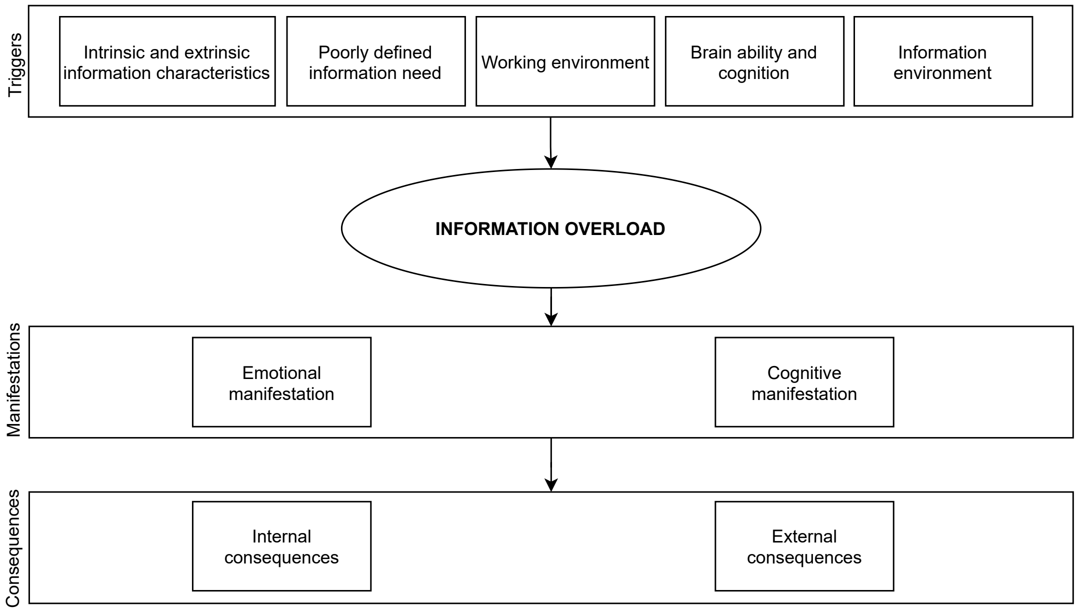


Figure 2.3: IO framework summarising the findings from the concept analysis

quality, diversity, novelty) and extrinsic characteristics are about the way the information is presented (i.e quantity of information, speed of delivery of the information, sources of information)

One important aspect of intrinsic information characteristics which can predict whether an individual will be subject to IO is the consistency of information, it was found that consistent information would induce lower levels of IO while inconsistent would lead to large levels of IO [81]. Related to the quality of information presented, we find that poorly structured information can lead to IO [82]: This finding was highlighted by Poghosyan [83] who had found that journalists were able to write better articles when they accessed well-structured feeds of information, hence this limits the risk of journalists facing IO and for the people reading the articles to avoid IO. The same finding applied in the context of organisations and business executives who were able to have an improved decision-making as as result of well-structured feeds of information [84]. Similarly, when looking at the quality of information, Keselman et al. [66] highlighted that the vagueness of terms when interacting with information is a cause of IO. While Lincoln [85] presented two additional triggers of IO: The relevance of information (see chapter 4 for more information about relevance) and diversity of information (i.e whether the information faced is complementing what was previously accessed).

One aspect related to the extrinsic characteristics of information was highlighted by Kao and Peng [86] who found that when relying on specialised book reviews systems rather than a general search engine (e.g Google) had a better time and were faster to find useful reviews. Another important aspect is the speed of information: Benselin and Ragsdell [87] highlighted that when information is delivered at high speed, individuals could understand that they were subject to IO and led them to feel restricted cognitively when attempting to process and reflect on the information accessed. Additional aspects of extrinsic information found in our dataset are: Quantity of information, speed of information, and sources (regardless of the relevance) [73, 88].

Poorly defined information need.

An information need is the initial step of the search process [89], hence a poorly defined information need means that the individual is taking on a search trail which will not adequately cover the targeted search domain and might lead the searcher to irrelevant information through poor precision of the search engine and in addition, the searcher will have to crawl a large corpus of information and actively filter it without any support [90]. Advanced IR systems which can achieve high precision would effectively allow less risk of falling under IO: In the context of organisations, in our dataset we find Simperl et al. [91] who highlighted the importance of understanding the information need of the employees of an organisation for better performance in their tasks.

Information environment.

Information can be of great quality and of adequate quantity through its relevance but one component which can still enable IO is the environment in which the information is present. The information environment represents the interface and organisation of the information on the mean from which the information is accessed [25, 60, 92, 93]. Previous work leveraging eye-tracking was able to show how the cognitive load of the user of an interface positively correlated with the increase of complexity of the interface [10]: The study by Wu et al. [10] showed how the increase in interface complexity also negatively impacted individuals performance when using LED manufacturing systems. In addition, an information environment should have a level of customisation to offer in order for the user to filter the information presented given their information need [94, 95, 96]: The poor design of filtering mechanisms leads in an acute level of complexity in managing information as highlighted by Yang and Albers [96] in the context of an online reputation management system. As a result of good quality filtering mechanisms, individuals can reduce the quantity of information they are exposed to as a result of only facing relevant information [94, 95, 96]. The inability to filter irrelevant information can exacerbate a poorly defined information need and hinder the progress of an individual on a search task which will result in limited learning [46]. Furthermore, there has been research conducted on whether systems that display information familiar to all members of a collaborative environment pushes cognitive biases, such as confirmation bias, in decision-making [97].

Working environment.

The working environment consists of the physical context in which individuals are when interacting with information, rather than the actual mean on which the information is accessed: The working environment represents the surroundings of the individuals. As such, when individuals are having to direct their information to additional external sources rather than only having to focus on the information of interest, this can lead to IO: Pushed information, particularly unsolicited [98], as well as the characteristics of the task to be completed result in IO. A task under time constraint which requires crawling an array of information was shown to lead to IO [99]. Additional task dimensions such as task difficulty and the structure of the task (i.e not having breaks throughout) were found to cause IO [45, 73, 100, 101, 102]

Brain ability and cognition.

We have alluded to this cause of IO throughout this work. Brain ability and cognition relates to the executive functions of the individuals, the executive functions are high-order cognitive processes which allow individuals to process information in their working memory, and direct their attention (executive function are further defined and discussed in chapter 4). This trigger is the result of the referencing to a cognitive overload and the allusion to a working memory saturation in our dataset [56, 61, 103].

In addition, Virkus et al. [56] alluded to individual difference based on the training of individuals: In their work, information science students were more prone to IO compared to informatics students. Virkus et al. [56] found that information science students believed that through their experience of databases and determining relevant sources for information, they were able to better cope with IO. This finding complements what we previously highlighted as part of the extrinsic information characteristics, i.e the sources of information play an important role in determining whether IO will occur or not [73, 88]. Finally, Virkus et al. [56] found that informatics students relied on their experience in using technology to overcome IO.

In our dataset, we identified the gender and age of the individuals interacting with information can determine whether individuals will find themselves subject to IO [61]:

In their meta-analysis, Shrivastav and Hiltz [61] highlighted research that has found that males were more prone to overload compared to females when reading scientific papers. In addition, Shrivastav and Hiltz [61] also found that younger individuals were less likely to feel overloaded compared to older-adults in an audio-visual task. Schmitt et al. [70] found that as individuals get older, they can lack strategies when exploring information and in turn this can result in IO.

Manifestations

When IO happens, it manifests itself within the indivdual in two ways: Emotionally and cognitively.

Emotional manifestation.

Emotionally when IO manifests itself, individuals are less prone to perform contemplative activities which can lead to feelings of sadness and/or depression [30, 48, 56, 104]. In our dataset, we found that a high perception of IO can lead to stress, which in turn results in emotions such as fear, and less confidence; consequently this leads to poor decision-making as we highlight in the consequences of IO [43]. Additional elements related to the emotional manifestations are fatigue, anxiety, and stress [42, 48, 105]. Hence, the emotional manifestations of IO involve the mental health of the individual [48].

Cognitive manifestation.

Cognitively, we define the manifestations of IO as being related to the individual limitations. Hence, when individuals reach the maximum of their working memory capacity, they are unable to further develop their knowledge [46]. IO leads them to a state of cognitive dissonance [106], they are unable to direct their attention and they feel burned out [104, 107]. This will lead to the internal consequences identified such as avoiding sources, or withdrawing from the search task [108]. In the context of education, Zhang [67] surveyed students who are users of an online collaborative environment: They found that students who were subject to IO were unable to concentrate and ignored information. Our dataset highlighted how individuals will lower their standards when under IO, hence this leads to a greater tolerance to error and acceptance of poor quality of work, as well as pessimistically approaching tasks [109].

Withdrawing from tasks and under-performing as a consequences of IO were identified as part of two types of consequences which we present next.

Consequences

The two types of consequences we identified are internal and external, these are driven by the manifestations the individuals suffer from as a result of IO.

Internal consequences.

At an individual level, IO will lead to limited learning which will slow the development of the indivdual's skills [110]: In the context of online healthcare information search, Swar et al. [90] found that due to the manifestation of IO, individuals withdrew from their search task, this led to limiting their learning. While similar findings were presented by Sabeeh and Ismail [71] in the context of organisations, venture capitalists subject to IO suffered from limited learning, as well as poor creativity in their problem-solving. As highlighted above, attention is tightly related to whether an individual will be under IO or not, but we found that attention can also be impacted when individuals suffer from IO resulting attention issues [26]. Additional internal consequences of IO are the poor well-being of the individuals [73, 111], poor creativity [31, 56, 71, 112], working memory saturation [30, 31], and a decrease in confidence when making a decision [113].

External consequences.

As we have seen, IO can lead to limited learning, hence this will effectively impact the individuals' external environment when using the information collected. The consequences on the external environment are the result of decision-making, a decision-making which is negatively impacted [112]: A poor decision-making can lead to financial loss [31], as well as human casualties [114]. One example of the latter is the 2010 US drone attack which targeted criminals but ended up killing 23 Afghan civilians. This was the result of the drone operator facing multiple feeds of information and eventually ignoring some [114]: This highlights the urgency in finding solutions to IO, solutions which can only be efficiently designed if the problem is well-defined. In addition to poor decision-making, IO can result in poor productivity, dropping out from tasks [90, 108], and hindered collaboration [23, 31].

Definition

Based on the main themes and groups identified above (Figure 2.3), we provide a definition for IO which has been published with this work [9]:

"Information overload is a negative psychological state in which individuals feel that they are receiving too much information, which hinders their ability to carry out their tasks. IO manifests itself through emotional and cognitive challenges and is most likely to happen through intrinsic and extrinsic information characteristics, poorly defined information needs, their working environment, brain ability and cognition, or the information environment. The emotional and cognitive manifestations result in both internal and external consequences. The internal consequences limit the ability of individuals to learn and inhibit their creativity. The external consequences impact the working environment of individuals. It translates into poor decision-making, which can result in both financial and human cost."

2.5 Discussion & Conclusion

In information science, David Bawden has paid a close look at the problem of information overload (IO), and his early work defined IO as the moment when information becomes a hindrance rather than help [20]. Yet, in 2020 in his subsequent work on IO, Bawden and Robinson [32] described IO as lacking consensus on a definition, and it has now grown to be associated with additional causes. A number of factors were identified as contributing to information overload: Among these factors are the quantity of information, the diversity, the complexity and the novelty of the information, pushed information, as well as factors related to individuals.

The work we presented in this chapter offers a definition and a framework (Figure 2.3) which covers all of the aforementioned causes of IO (RQ 1). Our framework (Figure 2.3) separates information characteristics into intrinsic and extrinsic information characteristics; pushed information falls under our working environment trigger;

and individual factors fall under our brain ability and cognition trigger.

We studied the various causes and consequences of IO in computer science and information science research based on the literature published between 2010 to September 2020 with the goal of providing a comprehensive definition of IO. We found that IO and cognitive overload (CO) were one and the same in computer & information science (RQ 2), and were able to identify main themes of IO which we incorporated into our main contribution which is an operationalised model enabling a better understanding of IO.

This chapter started by presenting one of the most common definitions of IO in information science, which defines IO as a phenomena that occurs when there is a huge volume of relevant and potentially valuable information; nevertheless, this information becomes an impediment rather than a benefit to an individual's learning. In addition, we also presented the definition of CO from cognitive science, which states that CO occurs when an individual's volume of information is too large, resulting in poor cognitive performance. However, we argue that both of these definitions are too narrow in the context of the present era. This work has clearly demonstrated that IO is about more than simply the quantity and relevance of information.

Finally, we presented a framework (Figure 2.3) that can be utilised to better understand and study IO in an information driven society. Our framework (Figure 2.3) established the triggers (i.e causes) which can be leveraged to create IO on individuals in a controlled environment. Being able to create IO in a controlled environment can allow to better study the individual's experience of IO and detect the instant IO occurs through its manifestations. These triggers and manifestations can help into the design of scalable models which can be used for the development of adaptive systems. Being able to study the triggers of IO in a controlled environment can also allow to determine their contribution in the state of IO, i.e are all triggers equal ? Additional work is necessary to evaluate whether these triggers can be assigned a weight based on their impact as well as the time it takes for those triggers to manifest IO internally. Some of the identified triggers such as the physical environment (working environment) in which individuals find themselves when performing a search task would be more difficult to reproduce for further studies.

In the future, we believe the research community should focus on identifying the potential different levels of IO, and how the individuals perceive IO differently based on their cognitive abilities. In order to perform such research, the identified triggers can be used to create IO and the consequences of IO can be used to evaluate the impact of IO (e.g performance on tasks, quality of decision making). Future work should also pay attention to the manifestations of IO to potentially design subjective measures which can allow to evaluate whether IO successfully happened, identify whether there are inter-individuals differences, and if there is a long lasting impact on the individuals, e.g how is attention impacted on the long term.

In this work, we contribute to the theoretical underpinnings of IO: We have stressed the importance of unravelling the various aspects that constitute IO, thus producing a definition and framework (Figure 2.3) which presents key aspects of IO. By identifying the triggers of IO, we can develop solutions but also effectively evaluate them by measuring how successful these solutions are based on how individuals react (manifestations) and how they perform in tasks requiring the use of the accessed information. A first principles approach allows to target specific components of systems, and evaluate those changes. While system improvements can be limited due to the progress of technology and developments of new algorithms, our work also highlights the importance of the executive functions of individuals when facing information: Supporting individuals based on their cognitive abilities should be prioritised by designing systems which can adapt to the individual or by attempting to improve the individual's executive functions. Based on the findings of this concept analysis, we design two studies: The first study looks at the triggers of IO that are related to the internal and external information characteristics (chapter 3), the second study examines the brain ability and cognition by looking at how attention, and working memory can be improved to tackle IO in a search task (chapter 4).

2.5.1 Chapter summary

In this chapter, we highlighted the lack of consensus regarding the definition of IO. We have seen that while in the literature, there is work looking at tackling it, no clear framework allows to scope these solutions and efficiently evaluate them. We collated our own dataset that was constituted of a total of 161 documents. This collection allowed us to define the concept of IO and present a framework to better study the concept in the future.

In the next chapter, we use our findings and design the first study to use identified triggers from our definition. The study's aim is to pave the way for further empirical work which can facilitate a better understanding of IO, exploring its dimensions, and identify opportunities to tackle it from a system-side view.

Chapter 3

Recreating Information Overload in the lab: A Twitter study

In this chapter, we went a step further in our attempt to better understand information overload (IO). We used our findings from the last chapter, and designed a study which allowed us to create IO in a controlled environment. Part of our definition from the previous chapter states that "IO manifests itself through emotional and cognitive challenges and is most likely to happen through intrinsic and extrinsic information characteristics", part of the extrinsic information characteristics are speed and quantity of information. This led us to design the following study with frequency (i.e manipulating the quantity and speed of information) as our trigger for IO. We collected log data throughout the study and based on the literature reviewed, we used mouse tracking to show how IO can be detected based on behavioural changes. Using the data collected, we developed classifiers which can predict whether a user was under IO or not, we also established the evolution of how IO affects individuals over time using clustering analysis complemented with self-reporting. To verify that our participants were under IO, we used subjective measures, in this case NASA-TLX, combined with self-reporting and semi-structured interviews to clearly establish that IO happened. The following chapter is structured as follows: First, we present a review of the literature (section 3.2) about information overload and social network sites, in addition, we review how past work looked at behavioural data to understand individuals' states. Following the literature review, we present the methodology (section 3.3) and our approach to answering the research questions. Then, we present our results (section 3.4). Finally, we discuss our findings and present the limitations of our work, as well as a conclusion (section 3.6).

3.1 Introduction

Information overload (IO) on social network sites (SNS) has seen a growing interest by the research community [25, 42, 62, 83, 115]. However, past work has heavily relied on theory, surveys and previously collected user data with limited information on how users interacted with a given social network, i.e the researchers would not know which posts or tweets a user actually read and therefore would have to make assumptions on how users interacted with the system [62].

With this in mind, in this chapter, we want to move past these limitations and create 'IO in the lab'. One important objective of this work is to pave the way for future research and to show that IO can be studied in the lab in a controlled environment, and we can build upon the findings of a study like this to contribute to the progressive disambiguation of the concept of IO. It also offers the research community as well as the industry indications on behavioural data which can be leveraged to detect IO.

While in the context of search tasks and computer and information science (CIS) in general, we find work that relied on user studies to measure cognitive load, but not with social network sites (SNS): This work heavily relied on dual tasks [116, 117, 118] and subjective measures [119, 120].

In order to measure how IO impacted the individuals' behaviour, studies leveraged implicit and explicit interactions of users [118] as well as eye-tracking [121]. Gwizdka [118] investigated cognitive workload manipulation of users while performing a search task and used a dual-task model: The primary task was the search task and the secondary task was the Stroop task [122]. They used reaction time on the secondary task as an objective measure of cognitive workload and a 5-point Likert scale to assess difficulty as a subjective measure of the cognitive workload (1: Very difficult to 5: very easy). While multi-tasking is a cause of IO [9], such a task would generate interaction data with the system which would not necessarily translate well to real-world scenarios as participants are moving between two tasks, i.e two separate interfaces. In addition, the objective measure relies on reaction time in the Stroop task, which is an uncommon task that individuals will find themselves doing while performing a search.

It is, therefore, necessary to perform more empirical studies in realistic environments to better understand how individuals react to IO: This would allow for a more robust understanding of IO in a real-world environment and how to control it. This would also lead to an improvement in the solutions developed.

In their attempt to use different modalities to detect the cognitive load of individuals, Chen et al. [123] highlighted the importance of developing intelligent systems which can adapt to the cognitive load of the users of such systems, and they presented a model which relied on speech and pen data to detect changes in cognitive load. However, such model for an adaptive information system is limited due to the nature of the modalities used: While speech can be collected when individuals are interacting with a voice assistant, combining this information with the use of a pen to draw circle shapes is difficult to translate in real world scenarios.

The use of neurophysiological devices to detect changes in individuals while they are interacting with IR systems can benefit the development of brain-computer interface devices and help draw new theories modeling the behaviour of users [124]. However, the use of external devices has an increased associated cost and can be inconvenient. A more straightforward approach which can lead to an objective assessment of IO is the implicit interaction of the users with the system, namely clicks, dwell time, and mouse movement. In the specific case of social network sites (SNS) and Twitter, we can track the like count, the tweets hovered on, and the position of the mouse across time.

Previous work in CIS looked at mouse movement and how unconscious mouse movement correlated with a high cognitive load: A study used data from the workload test set in the affective dataset [125] to investigate unconscious mouse movements, and they derived three outcome measures: Movement frequency, movement duration, movement position change (calculated using the difference between two mouse coordinates). The dataset provided mouse coordinates with their associated timestamps from a study that relied on a dual n-back task to create three levels of overload (low for dual 1-back, medium for dual 2-back, high for 3-back): The dual n-back task consisted of remembering an auditory and visual cue, and reporting when either matched using a mouse click. Therefore, performance did not rely on mouse movement. The dataset also included the NASA-TLX scores of each participant. An ANOVA test found that mouse movement frequency positively correlated with an increase in cognitive load while movement duration and movement position change showed no difference across conditions [126]. While such study provides insights into the use of mouse movement as an unobtrusive way to detect whether participants feel overloaded, the findings are limited as we do not know whether they translate well to a more realistic environment.

In this work, we use our previously defined framework of IO [9] to design a controlled user study in a highly realistic environment. We investigate the problem of IO in the context of SNS and put the participants under an IO and not IO condition. We chose Twitter as the SNS environment for our study, this is due to the dynamic nature of the social network, and how tweets can arrive and quickly disturb the structure of a feed. Twitter has also become a space for not only the general public but also journalists to discover unfolding news stories live, which is one of the basis for the design of our task scenario. The aim of this study is to test our findings from the concept analysis and establish the behavioural changes of individuals when under IO as they interact with SNS. If we achieve these objectives, we answer the following questions:

- **RQ 3**: What is the behaviour of individuals under information overload when interacting with an information retrieval system?
- **RQ 4**: What is the behaviour of individuals which are not under information overload when interacting with an information retrieval system?
- **RQ 5**: How is information overload perceived over time?
- **RQ 6**: What are the important features to develop a session-based information overload classifier?

To answer these questions, we use both objective and subjective measures. We use the participants' interaction data with the system as our objective measure, and IO manifestations and NASA-TLX as our subjective measures of IO. The interaction data consisted of implicit (hover on tweets, mouse tracking) and explicit (like) interactions. Past work has shown that the mouse can be used to track people's attention and experience [17] when they are browsing web pages, we rely on this finding and define various mouse related outcome measures to see how the participants' mouse related outcome measures changed across the IO and not IO condition. We create IO by using frequency of tweets as our trigger. The specific tweet frequency for the IO and not IO condition was identified after conducting pilot studies. We specify the parameters used in the methodology (subsection 3.3.3).

3.2 Background

In this section, we first review the literature on information overload (IO) in social networks and we complement it with additional literature related to the technologies involved in better modelling individuals' behaviour when using information systems: This review helps the reader better understand our motivation to use mouse-tracking as a viable substitute for more modern approaches such as neurophysiological devices.

3.2.1 Information overload in social network sites

Social network sites (SNS) consist of pieces of information generated by individual users. SNS can have a positive impact on individuals and society [127, 128, 129]. Erskine et al. [128] have found that the use of Twitter by academic journals to promote research positively correlates with an increase in readership and in citations. In addition to the facilitation of the dissemination of knowledge, SNS have such impact that the economy of a country can be affected when they are not accessible: Anyim [129] highlighted the negative impact of a potential Twitter ban on Nigeria's economy as small businesses rely on it for marketing purposes. This same ability to improve lives can also be manipulated nefariously and oftentimes can be used to spread propaganda and to shape narratives [130, 131].

Helmus and Bodine-Baron [130] show that ISIS' opponents constitute the majority on Twitter, yet they still have less of an impact, this is often described as the silent majority phenomenon [132]: They posted less on Twitter while ISIS supporters shared more content, and were better organised in sharing information by having coordinated attacks and using specific keywords such as "breaking news". This large quantity of information can create an IO state in individuals and lead to the perception that ISIS' supporters are more present, which in turn can impact individuals' emotional states by causing distress.

On SNS, information that is accessed is therefore not subject to editorial changes and can consist of anything a user wants to share: Albeit efforts to better control the content, users are still able to overcome these filters through obfuscation, i.e: neutralising any automated linguistic analysis by using alternative terms [133]. It is therefore very easy for a user to be flooded with a stream of information which can be irrelevant, harmful, or consisting of misinformation i.e fake news [115]. In a study looking at the reasons behind the spread of misinformation, it was found that IO positively correlates with "information strain" which, in turn positively correlates with the spread of fake news from SNS users [134]. Bermes [134] defines information strain as the feeling of fatigue, anxiety, and invasion of information [135, 136, 137]. These characteristics of information concur with the manifestations from our framework of IO [9]. Another manifestation we identified is attention deficit which was investigated by Hodas and Lerman [138]: While investigating the importance of tweets' visibility and the users' attention in the spread of tweets, they found that the traditional first-in-first-out queue structure for timelines played a role in whether tweets would be retweeted (reshared) by users'. It required a higher cognitive effort to seek older tweets, so these tweets were simply lost. In addition, it was found that users with a high following count have their attention split across their followings, which limits their ability to keep up with tweets that might be relevant but have been lost in a stream of new tweets.

To tackle the flood of posts, SNS started offering alternative views of users' timelines. Twitter, for example, offers a timeline that displays tweets using their own deep learning-based ranking algorithm [139]. Their work has shown that it increased engagement with tweets as well as engagement on Twitter. These last two points indicate that this new approach to displaying tweets reduced the IO which can be the result of a chronological timeline. This approach to solving the issue of large quantity of information and relevance was already being tackled by traditional search engines which used recommender systems to help filter information and offer more relevant information to the searcher based on the search trail of previous users with similar information needs. While the results of this ranking algorithm have been beneficial for the platform's engagement, some users still prefer seeing tweets in chronological order [140]. The reason behind this discontent is due to the algorithmic biases these non-chronological timelines can subject users to. It was found that they were limiting users' exposure to their friends' tweets [141], limiting exposure to tweets from under-represented communities [142] but can also create an echo chamber [143, 144].

The IO present on SNS has also impacted the consumption of news. Holton and Chyi [69] have shown that the consumption of news on Facebook was positively correlated with what they call "news overload".

Other SNS such as Tiktok and Instagram have developed solutions for users to control their usage of their platform. For example, Tiktok was offering users the option to receive reminders to take breaks. Later on, through another update, the company went further and now allows users to decide if they want to have enforced breaks after some time: These features are part of what they describe as an effort to boost "wellbeing" [145].

3.2.2 Neurophysiological insights in CIS

In recent years, there have been novel approaches aimed to better understand how people interact with social networks sites (SNS) and search engines using what is described as neurophysiological tools [146] such as eye tracking [147, 148], galvanic skin response devices [149], and at a more visceral level, brain imaging techniques have been employed to better identify the brain regions activated during the search process [124]. However, it is extremely difficult to use these findings in real world scenarios. In the case of eye tracking, devices' quality varies and they can all be configured in different ways, there needs to be a calibration step before they are used [150]: the distance between the user and the screen varies from one eye-tracker to another, as well as the frequency of operation, and wearable vs remote eye-tracker. While in the case of leveraging brain activity to establish individuals' states and develop neuro-feedback systems, there is still a lot of work to be done before we can reach a mainstream deployment of such devices [151].

In addition to the research done with eye tracking and brain imaging, there has been work which looked into identifying individuals' emotions through facial feedback while interacting with computers in order to have machines develop social awareness and improve their interaction with humans [152].

An unobtrusive approach to modelling the individuals' state is mouse tracking. As eye tracking has shown promising results in modelling individuals' states, research has tried to use the mouse as a proxy for the gaze. In the next sections, we present work done with eye tracking in CIS and look at how it was used to measure cognitive load in CIS. Subsequently, we venture into research that used mouse tracking: We review work which looked at the mouse-gaze relationship and then present studies which used mouse tracking to understand the individuals' behaviour.

3.2.2.1 Eye tracking

Eye tracking is a method designed to have real-time tracking of eye movement, blinking as well as pupil diameter [150]. It has been explored as a tool in human-computer interaction and IR tasks to better understand various aspects of users' interactions [153, 154]. One important axiom behind the use of eye tracking to model individuals' behaviour in the case of information search, is that the eye movement can be used as a proxy for attention: This is based on the eye-mind hypothesis, "the eye remains fixated on a word as long as the word is being processed" [155]. This view on visual attention has been shown to not be that straightforward, as individuals can look at one area in the space while at the same time having access to periphery vision [156]. Additional work further explored the eye-mind hypothesis and found that when individuals' gaze is on a word, there is the time to encode the information as well as a further path that is associated with the specific task being performed. This latter limitation is often overlooked by researchers in CIS.

Previous work has looked at using eye tracking for implicit relevance feedback, more precisely they investigated how text document processing changes when individuals are looking at relevant and irrelevant documents [157]. They found that there are two types of behaviour: Reading for relevant documents and scanning for irrelevant documents. By leveraging these findings, and filtering the type of fixations and limiting them to reading fixations, Bhattacharya and Gwizdka [147] have shown that eye tracking can be used to show when a higher knowledge gain happens: In a search based study, they assessed participants' knowledge before and after performing the search tasks, as participants were searching they had to take notes and bookmark relevant web pages. To calculate knowledge gain, they used a free recall assessment (participants had to share what they know about the topic) and used two methods: Relative difference in the number of items entered across the pre-task and post-task assessment, and the mean-rank of nouns. They found that participants with a lower knowledge gain invested more effort in trying to read, as they had more fixations on the articles than on the search engine results page, they fixated longer in total, and in average had a higher probability of reading, and had longer reading sequences.

Eye tracking has also been used in the context of SNS to better model the behaviour of users: Work by Wan Adnan et al. [148] investigated the use of Facebook and found that across the various user interface components, the participants of their study looked first at their timeline and spent more time reading their friends' post. Similarly, still in the context of Facebook, Simko et al. [158] used eye tracking in a study investigating how misinformation spreads. In this latter work, the goal was to extract behavioural characteristics which can be used to develop automated systems which mimic humans when filtering fake news. One of their findings is that individuals who were more prone to believe fake news mostly focused on the timeline items and didn't click on the related link to investigate the news any further.

In the field of CIS, there have been attempts to measure cognitive load using eye tracking [121, 159, 160]. Chen et al. [121] designed a study that relied on participants' recall of basketball players positioning to asses how high cognitive load impacted individuals' behaviour with three levels of cognitive load (low, medium, high). They performed the analysis on low and medium cognitive load (high was discarded due to corrupted data): They found that under higher cognitive load there was an increase in fixation duration and decrease in fixation rate (fixations were considered from 200 milliseconds), in addition to an increase in blink latency and pupil size while the blink rate, saccade speed and size decreased. The findings regarding the fixations' related outcome measures fit well within other bodies of research outside of CIS, where it was shown that an increase in fixation duration positively correlated with an increase in cognitive load [161]. The study by Dalveren and Cagiltay [161] relied on the ability of the participants to recall previously but recently seen information which would involve short-term memory. In this study the researcher uses the term 'working memory' which in cognitive science is a type of short-term memory, we further discuss working memory in chapter 4. Hence, if we want to fit this work within our framework of IO [9], it would be part of the 'brain ability and cognition' trigger.

In a study involving a visual search task to modulate IO, Zagermann et al. [162] highlighted the value of eye tracking in measuring the cognitive load of users: They found that a higher number of fixations and saccades per second and a decreased number of saccades indicates higher cognitive load. While the decrease in the number of saccades matches the findings of Chen et al. [121], the number of fixations per second and saccades per second can not be compared to the analysis by Chen where they looked at the fixation duration and saccade rate on the whole time taken for recall rather than per seconds. In addition, while one task relied on recall, the other relied on what we describe as the 'information environment', therefore it is possible that different triggers lead to different reactions.

All this work we reviewed highlights the advantages of using a non-invasive neurophysiological technique to better model humans' behaviour. However, still, in the case of eye tracking, we need additional equipment which can be calibrated in different ways. As a potentially viable alternative that is less obtrusive, we look at mouse tracking.

3.2.2.2 Mouse tracking

In the field of IR, several researchers have explored the relationship between the gaze and the mouse cursor [163, 164, 165, 166].

While investigating the movement of the mouse and the eyes, Rodden et al. [163] asked participants to perform search tasks and they tracked their eye and mouse movement on the search engine results page: They found that they used the mouse when they identified relevant search results to click, they also found a strong coordination in the movements along the x and y axis as well.

Huang et al. [165] investigated when the mouse and the gaze were aligning. They designed search tasks and collected various implicit interaction data in addition to the mouse data, and used eye tracking to track the participants' gaze: They found that the search topic, the user, and the time spent on a search result (the distance between mouse and the gaze decreased as they spent more time on the results page) all influenced whether there was a gaze and cursor alignment. Another finding was they were able to better predict the gaze position by comparing it to the future mouse position: The mouse lagged behind the gaze. These results showed that it is not necessarily true that the gaze and mouse are aligned, but there is a close relationship between the two.

The work reviewed so far focused on analysing the eye-gaze relationship on search engine results pages, a linear, single column web page. A subsequent study by Navalpakkam et al. [166] looked at this relationship but on non-linear web pages, they remained in the context of search engine results pages but added an additional column to the results page (the additional knowledge panel which sometimes appears on Google, see Figure 3.1). They found that the eye-mouse alignment is strong on non-linear web pages. They concluded that while the mouse behaviour can vary within and between individuals (active mouse usage while scanning vs idle), the mouse can be used as a weak proxy for eye movement and henceforth a weak proxy for attention.

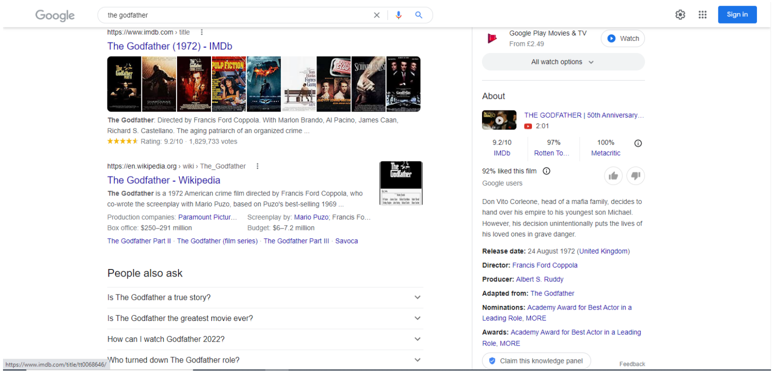


Figure 3.1: Search engine results page with additional modal

Recent work by Milisavljevic et al. [167] showed that eye movement and mouse movement have similar tendency overtime during the exploration of a web page which is in agreement with what was found on the search engine results page [165].

Research looking at the gaze and cursor alignment has shown promising results and would indicate that the cursor could be used to model users' attention as well as being a useful measure of cognitive load. Recent work went on to explore how cognitive load affects mouse movement [126, 168, 169].

Grimes and Valacich [168] designed a study where they modelled three levels of cognitive load using three slightly different tasks: The first level consisted of clicking on two different buttons depending on whether they saw the target number. The second level required comparing the current number displayed with previously seen and clicking on a button depending on whether the current number was smaller, or larger. The last task required a higher recall ability, the participants had to sum two previously seen numbers and see if it matched the currently displayed number. To ensure that the three levels were modelled they also used NASA-TLX to confirm the statistically significant load difference across the three tasks. They found that users under high cognitive load took a longer time to complete the tasks, had a larger travel distance, slower mouse speed, and higher number of direction changes. While these outcome measures are valuable for such a task as it required an element of speed, it is safe to assume that when participants had to do the mental recall and arithmetic, it would take them more time to move the mouse. It would also be difficult to assume that the individuals' behaviour would be the same on various cognitive load levels when browsing on a search task or when participants are crawling social network sites (SNS).

A similar and more recent study used a dual n-back task to represent three levels of cognitive load: Cha and Min [126] found that the unconscious mouse movement represented through the mouse movement frequency positively correlated with an increase in the cognitive load.

Finally, Navalpakkam and Churchill [17], in their effort to prove that mouse tracking can act as a proxy to eye tracking, they have shown that mouse tracking can be used to measure the individuals' attention and measure the overall experience of the user: They found that as the amount of distraction increased in a web page, the participants' mouse dwell time (no movement) increased.

Our work fits within this goal to leverage mouse tracking to model the individuals' experience while relying on an unobtrusive and scalable method: The importance of relying on implicit behavioural data remains important to support people when they are interacting with IR systems. The implicit feedback can then be used to develop intelligent adaptive systems. While the effort to move to more advanced techniques such as neurophysiological tools contributes to better understanding the users' behaviour, there is still a long way to go before reaching mainstream deployment. Hence, with the development of new IR systems and the new challenges associated with them, in this case, IO, we have to keep exploring methods within reach to improve the users' experience but also support their well-being. In this work, we used mouse tracking as our implicit behavioural data, and create new models to classify whether users were under IO or not. This approach offers a window for fast scalability to understand the users' state.

3.3 Methodology

In this work, due to the 2020 coronavirus pandemic, we had to run the user study remotely. It was conducted over the summer 2021, we were not able to access any additional device which could have complemented the findings of the study. Therefore, we focused on the explicit and implicit participants' interactions with the system as our signals for IO, and use semi-structured interviews to triangulate our results.

3.3.1 Ethics

Ethics approval was obtained from the department of Computer & Information Sciences' Ethics Committee, University of Strathclyde (application ID: 1477). Participants were given an information sheet (Appendix A.1), and a consent form (Appendix A.2) to be dated and signed. Prior to signing the consent form, they were asked to read the information sheet. In the information sheet, the participants were made aware of their freedom to withdraw from the study at any time and that their participation was voluntary. Participants provided informed consent at the start of the session once they read the information sheet. At the end of the study, all participants were given a debriefing form (Appendix A.3) with the contact details of the researchers involved in the study, the participants also received 7 GBP shopping voucher. All the collected data was anonymised: Each participant had a randomised user ID assigned when consenting to participate to the study. The data was stored in a secured location (password protected).

3.3.2 Participants

We recruited 32 participants through convenience sampling (advertising online and in-person). All participants were required to be healthy and fit the 18-40 years old age group, participants were also required to be fluent in English. In total, we had 17 male and 15 female participants.

We performed a post-hoc power analysis (Cohen's d=0.5, alpha=0.05, N=32) set with the Wilcoxon signed-rank test: We found that the power of the study was 0.8.

3.3.3 Study Design

This study used a within-subject design. The dependent variables were qualitative (questionnaires) and quantitative (interaction with the system) data. To avoid learning and fatigue effects, we utilised a Latin square design [170] for counterbalancing: Our participants were randomly split in two groups, one group (N=16) performed the information overload condition first and the second group (N=16) performed the information overload condition second. In each group, we randomised the order in which they performed the topics: One half performed topic 1 as the first topic, and the second half performed topic 1 as the second topic.

The independent variable was information overload, it was controlled through new tweets frequency (i.e amount of new tweets added at every update and speed of update). During the IO condition, participants received 7 tweets every 2 seconds while during the not IO condition they received 3 tweets every 7 seconds. The IO condition had a ramp-up period which lasted the first 135 seconds before overload inducement. The ramp-up period had two update speeds: In the first 45 seconds, 3 tweets appeared per 15 seconds then for the last 90 seconds of the ramp-up period they received 7 tweets per 10 seconds.

3.3.4 Experimental Procedure

The whole experiment execution took place over Zoom, an online videoconferencing tool. The participants were able to control the study lead's screen remotely. The system was not deployed for two reasons: We wanted to have control of the screen dimensions to allow us to have a consistent representation of the mouse movement and also the number of tweets appearing on screen, finally, we also wanted to reduce the latency related to the collection of system interaction data. At the start of the experiment, the participants had to go through the information sheet and sign the consent form. The experiment started with the entry questionnaire, followed by the pre-task questionnaire, and then they had to do the first task for 10 minutes. At the end of the first task, they filled out the post-task questionnaire and subsequently had to fill out the pre-task questionnaire for the second task. At the end of the second pre-task questionnaire, the participants started the second task and had again 10 minutes to monitor the Twitter feed and comment on the events as they were unfolding. The first and second task instructions were the same, however, the topic and condition order was randomised. Once the second task was over and the participants had completed the post-task questionnaire, they were provided with an exit questionnaire to complete.

Finally, to conclude the study, we asked participants questions as part of our semi-structured interviews.

3.3.5 Task description

In this experiment, the participants had to do the same task twice, one for each condition. It was meant to replicate the experience of being a journalist who is in a newsroom. The participants were asked to behave like journalists who had to narrate out loud the development of an event. Since we asked them to perform the same task twice, we had two different topics assigned. As journalists, the participants had to narrate the event as it unfolds, they had to like any tweet which might be useful to describe the situation and express it out loud whenever they felt overloaded. The scenario they were given is as follows: "You are a journalist and are part of a newsroom in charge of following the latest events through Twitter. During the next 10 minutes, you must follow the evolution of an event and keep your colleagues in the room updated. It means that you will have to narrate the development as they are happening on social media, and you will be required to share if you feel overloaded by saying it out loud. You should like (click on the heart) tweets which you find are useful to succeed at the task. At the end of the 10 minutes, you will be writing a summary of what happened for your colleagues, it will need to contain key details about the event."

3.3.6 Social network system

The Twitter clone was developed to offer us the ability to control the tweets presented to the users as well as various intrinsic elements such as the speed the timeline refreshes and the number of new tweets displayed at every refresh. It also allowed us to collect users' behavioural data: tweets hovered on, tweets liked, and mouse coordinates.

3.3.7 Twitter clone

3.3.7.1 Architecture

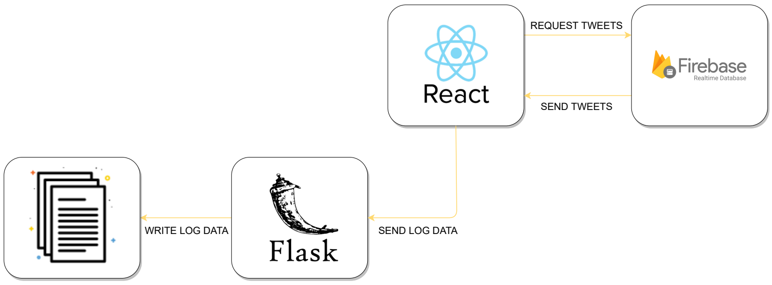


Figure 3.2: Twitter clone architecture

As our database, we used the Google's Firebase 'Realtime Database' which hosted our selection of tweets in a JSON format. Our frontend was developed using React and the backend was written in Python with a local server powered by the Flask framework. The React app interacted directly with the database and retrieved the tweets through API calls at every refresh. In parallel, the React app had a series of functions that collected real-time data based on the participants' interaction with the system's interface. The data collected was sent through a series of API calls to the backend which wrote directly data in locally stored comma-separated files. The files' names were formatted with the participants' ID and the task number (1 or 2).
The code is available at `https://github.com/aminecs/twitter`

3.3.7.2 User interface

In order to ensure that we had an interface that was similar to Twitter, we had to break down the various components that are a part of the Twitter timeline page, including the same colour scheme as found on the Twitter platform in our CSS files (figure 3.3). We removed new followings recommendations and trends to limit the distraction of the system's users, and we also parsed all the tweets as such they did not provide hyperlinks and all URLs appeared as text that they were not able to click on. The latter was an additional precaution as we had still told the participants to not leave the Twitter interface. In our Twitter clone only the main feed page was available and the like feature, all the other features/buttons were disabled/not implemented.

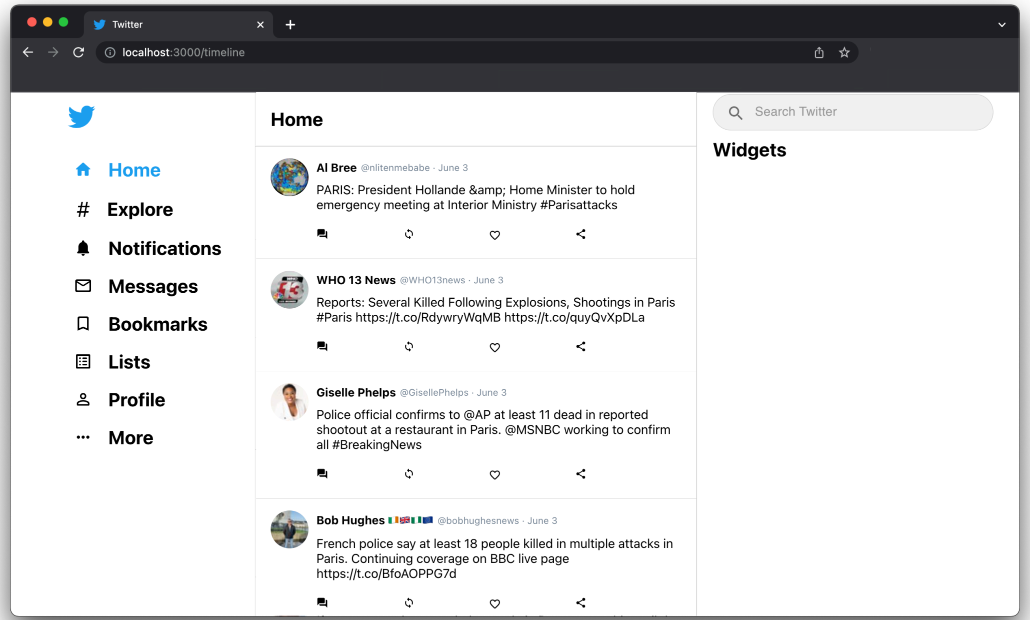


Figure 3.3: Screen of the built Twitter clone

3.3.8 Topics

We chose two topics from the "Twitter event datasets (2012-2016)" [171]: A dataset that contains 30 different datasets related to events which occurred between 2012 and 2016.

Due to experiment limitations, we had to choose two topics that had a linear and quick development. As such, we chose the Boston marathon bombing from 2013 [172] and the Paris terrorist attack from November 2015 [173]. While the events chosen are very tragic, due to our criteria of selection we were limited by the options presented by the Twitter dataset and had to include two tragic incidents in order to also avoid having a potential random effect if the events were of an emotionally different nature.

For both topics, the tweets were shown chronologically and started from the first reports regarding each event.

3.3.9 Questionnaires

At the beginning of the experiment, every participant filled out an entry questionnaire (Appendix A.4) which allowed us to collect demographic data, and learn more about their social network use. At the end of the entry questionnaire, they were given the task description and a pre-task perception questionnaire. Before each task, they had to fill out a short pre-task questionnaire (Appendix A.5) allowing us to know how familiar they were with the topic. At the end of each task, we asked them to fill out a post-task questionnaire (Appendix A.6) which consisted of asking them to summarise the event they just learned about, their perception, and the NASA-TLX form. At the end of the experiment, they were given an exit questionnaire (Appendix A.7) which collected information about how similar they thought the system was to the actual social network Twitter and which topic they preferred. In the exit questionnaire, we also asked the participants for any additional feedback.

3.3.10 Semi-structured interviews

In order to complement our findings and ensure that we induced IO, we conducted semi-structured interviews at the end of the study, after the exit questionnaire. The advantage of this approach is that it allowed us to triangulate our quantitative and qualitative data. The use of mixed methods was expected to strengthen our findings. While structured interviews follow a clear set of questions needing an answer, and unstructured interviews follow an open conversation. We used semi-structured interviews as they gave us the advantage of having questions answered but also allowed us to explore additional dimensions of the study that might be raised by the participants. The goal was to use semi-structured interviews so as to confirm that IO happened and complement the qualitative and quantitative data.

3.3.10.1 Procedure

The interviews were not recorded but we typed direct quotes during the interviews and our responses were collated onto a Microsoft Excel spreadsheet. The interviews were not recorded in order to ensure the participants spoke freely, and we did not want to add additional cognitive load as they were already put under IO, and they might have felt pressure when answering our questions. At the end of the interviews, participants were asked to confirm the statements collected. A set of questions were defined to allow the exploration of specific areas but the interviewee was left to speak freely:

- How would describe your interaction with each event?
- Did you adopt any specific strategy?

If they responded positively to the last question, we asked an additional question:

- Which strategy did you adopt?

To analyse our semi-structured interviews we used quantitative content analysis [174], we defined our codes based on the findings from our first study: We defined emotional manifestations of IO as one, failure to filter as a consequence (given our scenario), and one code for strategies. The main researcher and an external individual to the research performed the coding. Prior to the coding, the external coder was familiarised with our first study so they could understand the emotional manifestations and consequences of IO. The coding was done using the NVIVO software which allowed us to calculate Cohen's Kappa coefficient, a statistical measure for inter-coder reliability, it is used to compare the coding by two or more coders [175].

3.3.11 Notation

In our analysis, we denoted the IO condition as IO_1 and the not IO condition as IO_0. When IO_1 was done first, we represented that with 'n' and when IO_0 was done first, we used 'i': In some figures and tables, we will explicitly specify which order is represented.

3.3.12 Statistical analysis

In this section, we introduce the methods used for the analysis of our data in this study. Our data analysis was mostly performed using the programming language Python 3, and we used the library Scipy [176] for the various statistical tests used.

Ahead of performing any statistical significance test, we tested our data for normality using the Shapiro-Wilk test [177].

With normally distributed data, since we had a within-subject design, we could use the pairwise t-test. For non-normally distributed data, we used Wilcoxon signed-rank test. For our results, when testing for significance, unless explicitly stated otherwise, we used the Wilcoxon test as most of our data was not normally distributed.

3.3.13 Machine learning

Since our analysis will make use of traditional machine learning (ML) techniques. This section introduces some useful concepts of ML which will be of help in understanding our analysis. There are four types of machine learning categories: Supervised learning, unsupervised learning, semi-supervised learning [178], and reinforcement learning [179]. In this body of work, we will only discuss supervised learning and unsupervised learning as these are the ones we will use for our analysis. We use machine learning to classify our sessions whether a participant was under IO or not and to determine whether there are phases of IO based on the IO reporting of our participants.

In supervised learning, there are two types of problems: Regression and classification. We will not be discussing regression problems further, as we have a classification problem. Classification problems are about predicting two or more classes. In our case, we have two classes IO_1 and IO_0. In addition, we will use unsupervised learning as we have a clustering problem since we are trying to determine phases of IO based on unlabeled data (i.e data that does not have a class name associated with each given observation).

Firstly, we introduce supervised learning: a type of ML and present the various models we used to classify our data. Subsequently, we present unsupervised learning and the K-means clustering method we used.

3.3.13.1 Supervised learning

Supervised learning consists of training a model with a fully labeled dataset, i.e each sample in our training data is labeled accordingly and a target function is created.

For each sample, we have the associated features and a training phase, the models are trained to be able to predict which class these features represent.

In our case, we had two classes (i.e IO and not IO). This falls under a binary classification problem, and our features were the outcome measures we identified through the analysis. We used five types of popular models for our classification problem: Logistic regression, support-vector machine (SVM), decision tree, random forest, and XGBoost.

Classification algorithms aim to draw a decision boundary between our points, it creates regions within our plane and each region represents a class. This is the case for both logistic regression, and SVM [180].

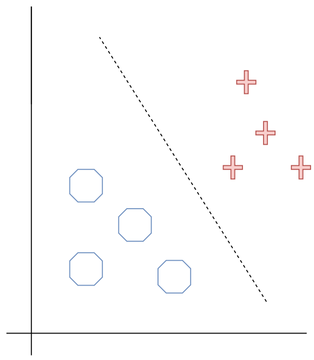


Figure 3.4: Visualisation of binary classification. The dotted line is meant to separate two classes.

In our case, we would have two regions since it is a binary classification problem (Figure 3.4). For example in the case of the SVM, the aim of the algorithm is to maximise the distance between the decision boundary and the points of the different classes [180].

Random forest, and XGBoost are ensemble models which are based on decision trees. Decision trees aim to classify points based on estimated cutoff values for the features. Decision trees shine when features interact with each other, however they struggle when there are slight variations in our data [181]. Random forest are a combination of decision trees. The difference between random forest and XGBoost is that for XGBoost the trees are built additively and each tree is meant to improve on the weaknesses of the previous tree [182]. Hence the term gradient "boosting" to describe XGBoost, the subsequent trees improve the model ("boost").

All of these algorithms fall under the category of discriminative models, these models are better when there are outliers in the data and usually perform better compared to generative models which are often used when there is unlabeled data [183]. To find out more about how each model works, see Molnar [184].

3.3.13.2 Unsupervised learning

Unsupervised learning differs from supervised learning as it does not need labels for the data. It tries to learn patterns from one or more features and separates the data into two or more classes. In our case, since we based our IO phases on the IO reporting data, we only had one variable/feature, therefore we needed a univariate clustering algorithm: One that is popular is K-means clustering. K-means aims to maximise the distance between the clusters and reduce the distance between the points that are part of a cluster [185].

Based on prior work, we found that K-means clustering performs better than other models for univariate clustering on small datasets [185].

3.3.13.3 Evaluation

We used two methods of evaluation for our classification models: Accuracy and F1-score. We report the F1-score rather than precision and recall individually, as we wanted to have a good balance across the two hence the F1-score was more indicative of how our models performed on both.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{3.1}$$

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.2}$$

Precision and recall are defined as:

$$Precision = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives + number\ of\ false\ positives} \tag{3.3}$$

$$Recall = \frac{Number\ of\ true\ positives}{Number\ of\ true\ positives + Number\ of\ false\ negatives} \tag{3.4}$$

To evaluate the K value that we chose from the K-means algorithm, we used the silhouette score [186] which allows us to measure the quality of our cluster and offers a visual aid to compare the score for various K values.

3.3.13.4 ML approach

In our analysis, we extracted mouse movement related data as our implicit behavioural data and used its features to build classification models. All our models were from Scikit-learn, a machine learning library in Python [187].

We built a pipeline using the function provided `make_pipeline()`. We passed two arguments to this function, the first one was the `StandardScaler()` which helps scale our feature and the second argument was the model we choose.

In addition to the building stage of our models, we also performed the evaluation on two metrics: Accuracy and F1 score.

Since we performed our user study on 32 participants, we had a total of 64 sessions. Therefore, our classification was performed on a small dataset. However, we did not have a class imbalance issue since we had 32 data samples for each class (IO and not IO). To overcome the issue of the size of the dataset and avoid overfitting, we used K-fold cross-validation. To implement K-fold cross-validation, we applied `cross_validate()` from Scikit-learn: This function allowed us to pass our model, our features and labels, the number of folds (k) the scoring functions and it took care of returning the model's performance across the different evaluation functions (in this case accuracy function and F-1 score). We decided to build the following models: Logistic regression, SVM, decision tree, random forest, and XGBoost. In addition to testing these models individually, we used `VotingClassifier` from Scikit-learn as an ensemble method, which allowed to combine the results from different models to get an optimal prediction.

While the logistic regression model and the support-vector machine were implemented without any additional parameter fine-tuning. The decision tree and random forest models had their maximum depth set to four, since we did not have a large dataset not specifying a maximum depth could lead to complex models that would result in overfitting [188]. The XGBoost [182] model had the number of estimators (i.e number of trees) set to 5, the objective parameter was set to 'binary:logistic', and the evaluation metric was set to logistic loss.

We also used our best-performing models for our data in an ensemble modelling approach, where we used the voting classifier with a soft voting policy, and 3 and 1 for the weights (3 for logistic regression, and 1 for support-vector machine).

We used the R programming language to perform the K-means clustering, and the CKMeans library for 1-dimensional clustering and to create our plot of the clusters [189], the `silhouette` function employed was from the R cluster package.

3.3.14 Mouse movement outcome measures

Time spent on the feed/timeline area

To calculate the time spent on the timeline, we first calculated the timestamp difference of every two consecutive events. Then, we identified the coordinates of the rectangle that represent the timeline and afterwards calculated the total time spent in the timeline rectangle.

Total travel distance

To calculate the total travel distance, we computed the difference of (x,y) coordinates between two consecutive events.

Screen regions explored

To define screen regions, we chose 20px by 20px: This choice was made heuristically as we found that 20px by 20px captured the action button of the tweets (response, retweet, like) for the screen format we had (2560×1600).

Fixations

In this work, when we talk about mouse fixations, we mean the moment in which a mouse did not move. As presented in our background, the mouse can act as a weak proxy for eye tracking [17]. Hence, we borrow from the literature on eye fixations [190] and define three types of fixations as presented by Velichkovsky et al. [190]: Express fixations which are the ones that last between 150 milliseconds and 300 milliseconds, modal fixations are the ones that last between 300 milliseconds and 600 milliseconds, and cognitive fixations are the ones which last between 600 milliseconds and 2 seconds. We discard fixations that are less than 150 milliseconds and fixations that last over 2 seconds to calculate the mean fixation time.

3.4 Results

In this section, we first report descriptive results: participants' demographic data, use of SNS, their experience with our Twitter clone, and their familiarity with the topics. We then examine whether we successfully were able to put our participants under IO, to evaluate this we use two approaches: The first approach looks at the number of times participants reported they were under IO in each condition (subsection 3.4.2), and the second approach looks at the results from the NASA-TLX questionnaires which were filled after each condition (subsection 3.4.3). To complement these two approaches, we report the results based on the manifestations from chapter 2 as an indicator that IO happened (subsection 3.4.4) and the topics' preference.

Subsequently, we look at the participants' behavioural data from two dimensions (subsection 3.4.7): The first one consists of the data generated from the interaction of the users with the system's components (i.e tweets in this case), and the second consists of the mouse movement.

Based on the IO reporting across the IO condition, we perform an univariate clustering analysis using an unsupervised learning method to identify the phases of IO. Once the phases are identified, we look at how the behavioural data varied across the IO condition and not IO condition.

Finally, we investigate the potential of using the behavioural data generated through the interaction with the system to design a classifier which can predict whether the user of a social network, in this context Twitter, was under IO or not.

Finally, we look at our semi-structured interviews to cross-validate our findings.

3.4.1 Descriptive results

3.4.1.1 Participants' education level

We find that across our 32 participants, 37.5 % were either in the process of completing a Ph.D. or were in the process of completing their first degree (figure 3.5).

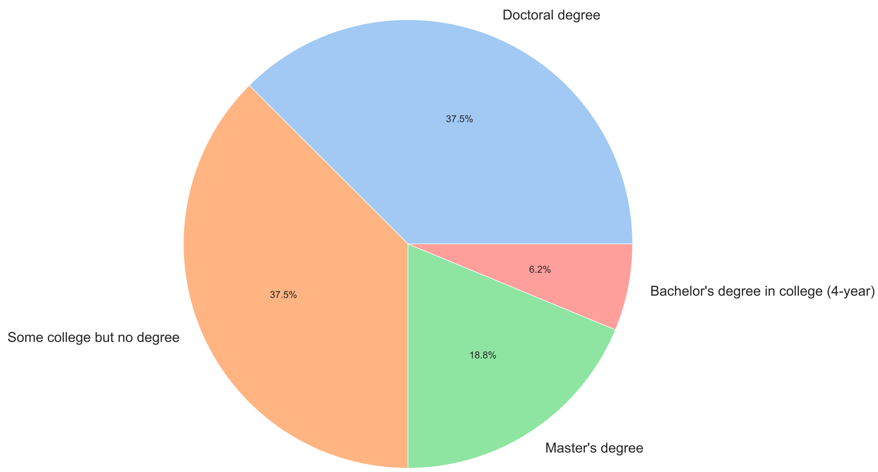
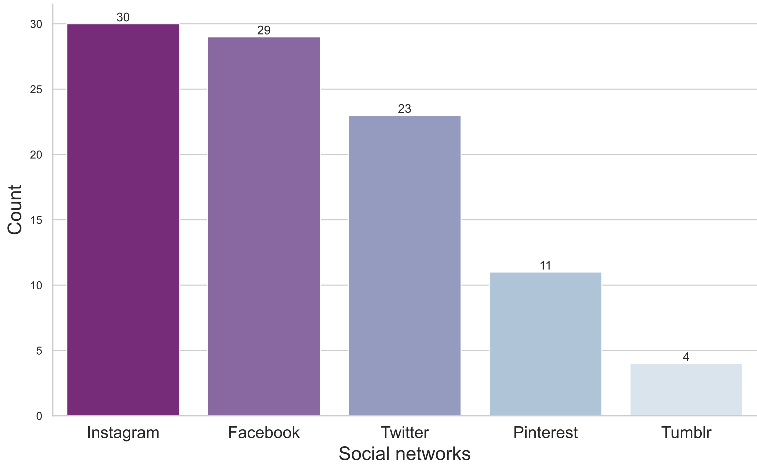


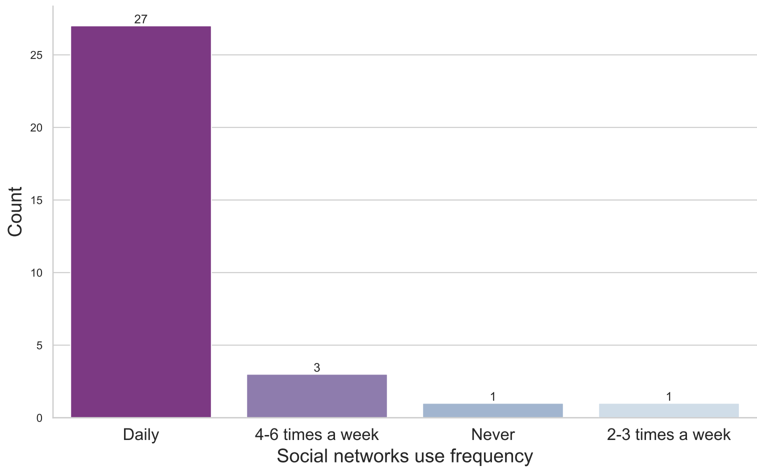
Figure 3.5: Education levels across participants

3.4.1.2 Social networks use

When looking at the responses regarding the SNS use of our participants, 27 reported that they used social networks every day. Within our group of participants, the most widely used social network is Instagram ($N = 30$) followed by Facebook ($N = 29$) then Twitter ($N = 23$) (figure 3.6a).



(a) Social networks used



(b) Social networks' frequency of use

Figure 3.6: Social networks used by the participants

While most of our participants shared that they used social networks every day (N = 27, figure 3.6b), 12 of them only opened Twitter "about once or twice a month", and 10 "never". The rest of the participants used it at least once a week (figure 3.7).

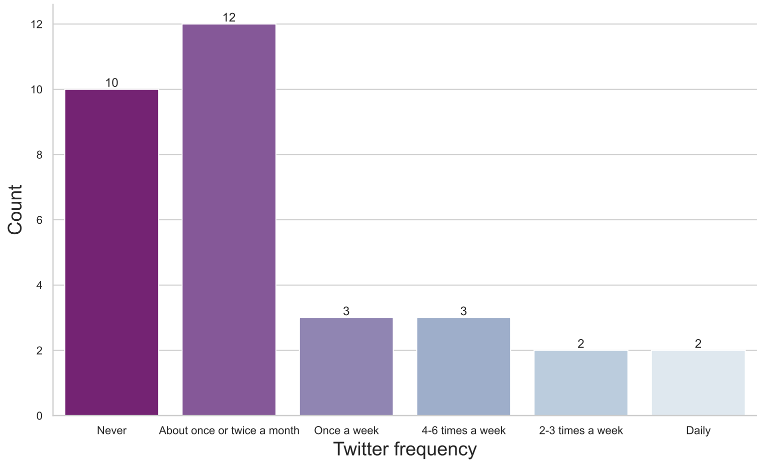


Figure 3.7: Twitter frequency use

3.4.1.3 Twitter clone similarity

At the end of the experiment, through a 5-point Likert scale (*"Extremely dissimilar"/"Dissimilar"/"Slightly similar"/"Similar"/"Extremely similar"*), we asked the participants to report how similar was the interface to the actual Twitter interface. We find that the majority thought it was similar (N = 15) and extremely similar (N = 13).

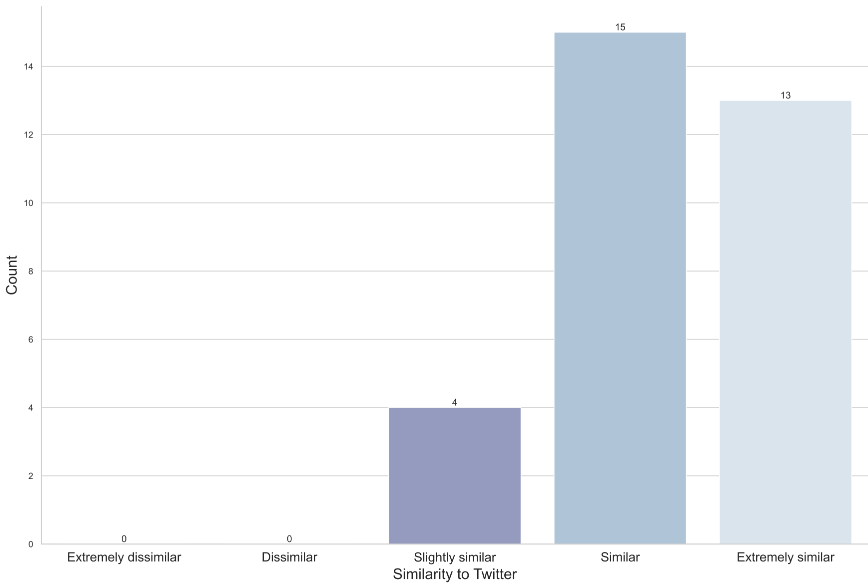


Figure 3.8: Twitter UI similarity

3.4.1.4 Topic familiarity

Out of our 32 participants, we find that 8 participants were unfamiliar with the Paris terrorist attack, and 10 participants were unfamiliar with the Boston Marathon bombing. The majority was at least slightly familiar to moderately familiar with our topics, with more participants slightly familiar with the Paris terrorist attack ($N = 11$) and more participants somewhat familiar with the Boston marathon attack ($N = 12$). Only 1 participant in each topic was extremely familiar (figure 3.9).

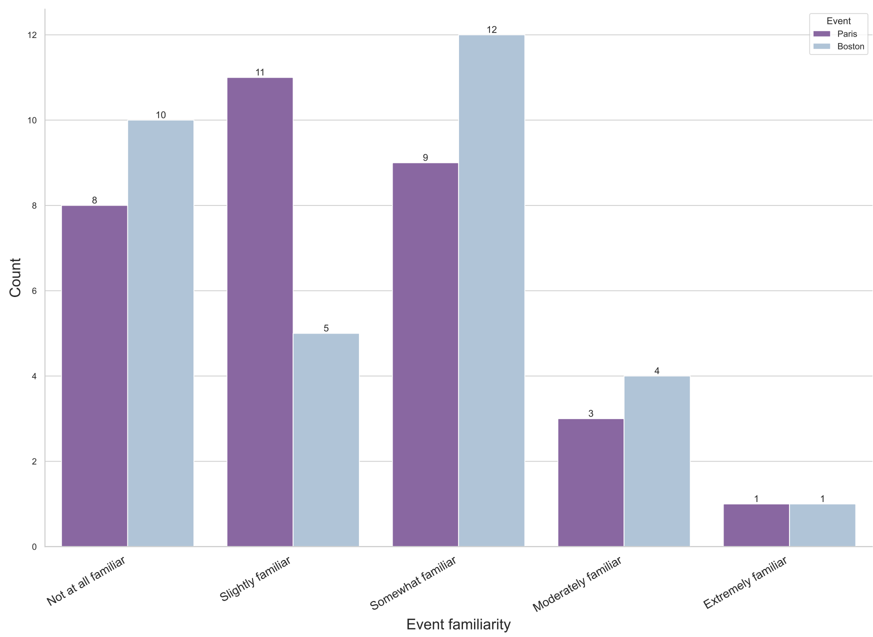


Figure 3.9: Topics familiarity

These results indicated that while most of our participants were aware of these two events, their lack of knowledge allowed us to create a loosely realistic experience with Twitter where they had to collect and filter out facts related to the events.

One explanation behind the lack of familiarity with these events is that our participants were recruited within a 18 to 40 years old age range, and since most were students (Figure 3.5), they were likely too young to know about them.

3.4.2 IO reporting

During our study, we asked our participants to share out loud when they felt overloaded. We found that there were more participants reporting IO under IO_1 than under IO_0 ($p < .001$). We can see in Figure 3.10 under the IO_1 condition ($Mdn = 6$, $IQR = 5.5$) participants reported the feeling of overload more than in the IO_0 condition ($Mdn = 0$, $IQR = 1$).

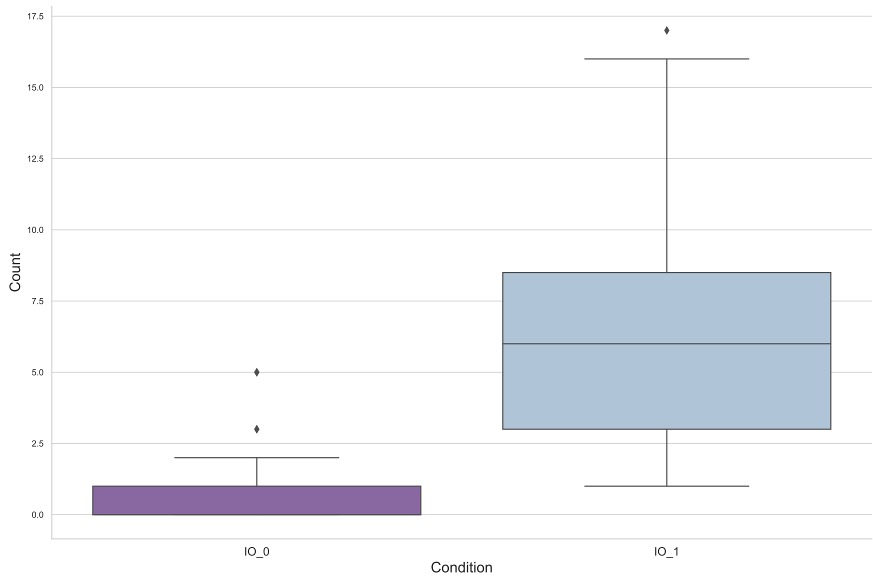


Figure 3.10: IO reporting during the full sessions across the IO_0 and IO_1 conditions.

In addition, we looked at the IO reporting given the order in which the conditions were attributed. We represented when participants did IO_1 first followed by IO_0 as 'n' and when participants had IO_0 first followed by IO_1 as 'i' (Figure 3.11).

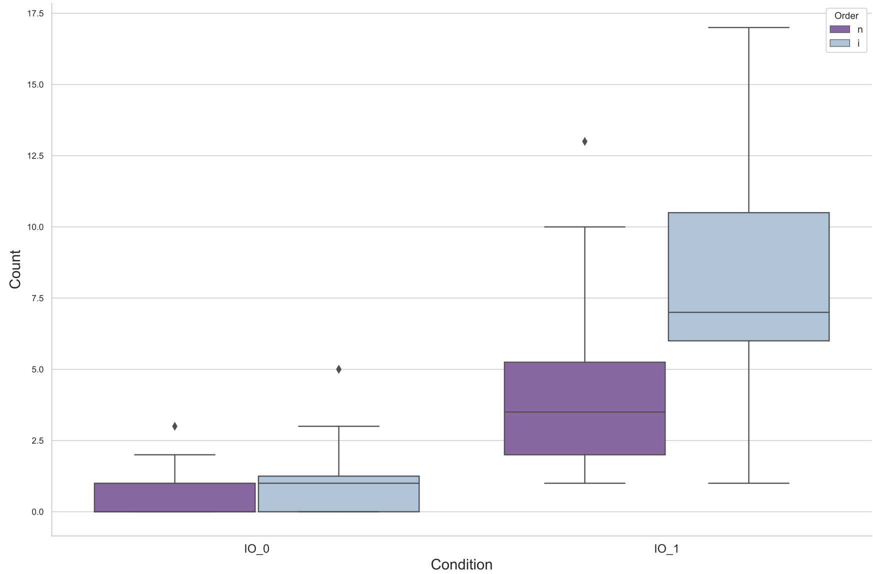


Figure 3.11: IO reporting during the full sessions across the IO_0 and IO_1 conditions when split according to the order in which the conditions were attributed. The order 'n' represents when IO_0 was done first, and the order 'i' represents when IO_1 was done first.

In the case of order 'n', more participants reported IO ($p < .001$) under IO_1 ($Mdn = 3.5$, $IQR = 3.25$) than under IO_0 ($Mdn = 0$, $IQR = 1$). Similarly when participants performed IO_0 first followed by IO_1, participants still reported IO more ($p < .001$) under IO_1 ($Mdn = 7$,$IQR = 4.5$) compared to IO_0 ($Mdn = 1$,$IQR = 1.25$).

Finally, we look at the evolution of IO reporting over time: We find that IO was reported 10 times by eight different participants before the 135th second (i.e time at which the ramp-up period ended). Out of the eight participants, six were performing the IO condition as their second task. One explanation regarding the participants who reported IO prior to the end of the ramp-up period is that they were feeling tired already as they had just completed the not IO condition. Regarding the two participants who were performing IO_1 as their first task: One participant reported IO at second 68, and the second participant reported IO at second 69 and 78. This reporting happened during the ramp-up period, exactly during the last phase (which starts at the second 45) where tweets were appearing at a slightly faster pace.

In Figure 3.12, we see that the majority of our IO reporting happens just after 135

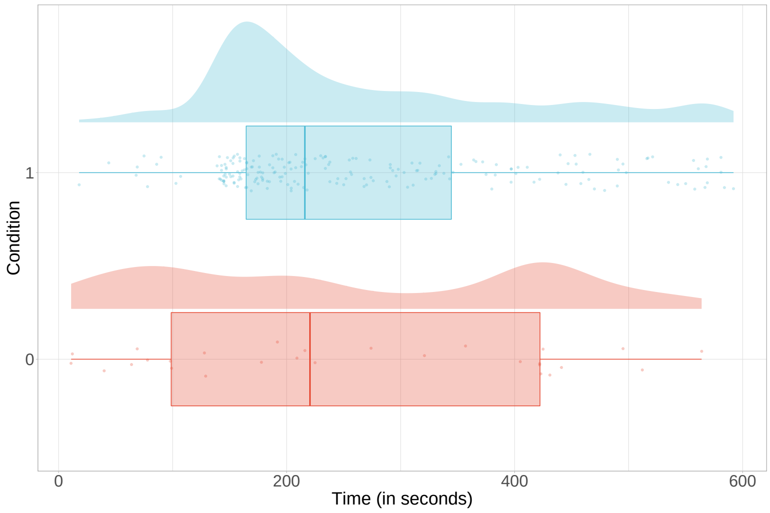


Figure 3.12: IO reporting across IO_1 and IO_0

seconds which is the time our ramp-up period ends and where our system uses the frequency set to overload the participants.

We plot the distribution of the IO reporting across conditions in Figure 3.13, and we can see that most values are around 135 seconds followed by a drop and an even distribution all the way to the end of the task.

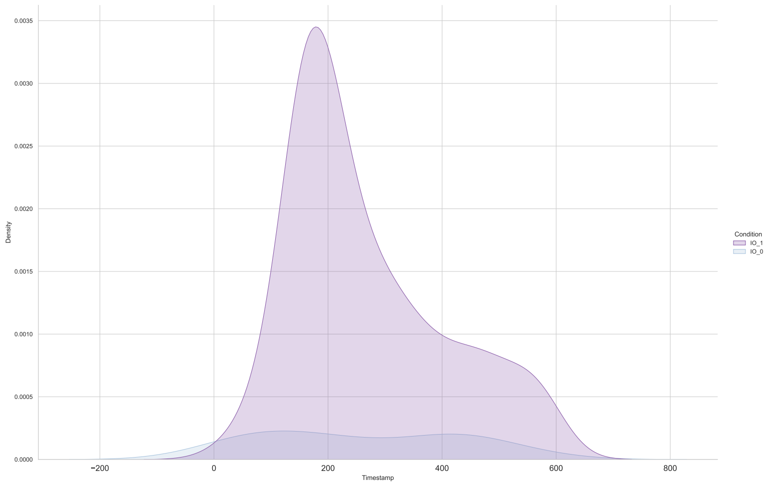


Figure 3.13: IO reporting distribution across IO_1 and IO_0

3.4.3 NASA-TLX

In order to measure the cognitive workload of our participants, we used the NASA-TLX questionnaire [191]. NASA-TLX is a subjective measure, it consists of 6 subscales: *Mental demand, physical demand, temporal demand, performance, effort, frustration.* Each subscale consists of a 21-point grade (min=0, max=20).

For the sake of our experiment, we dropped the physical subscale as our experiment did not require physical effort from the participants. In addition, we broke down the last component of the frustration subscale into five subscales since some of those components were identified as part of the manifestations related to IO from the previous chapter, and the question was identified as overloaded during pilot studies.

For each subscale, we calculated a Raw score which amounts to a percentage. Regarding the performance subscale, we reversed the score ("How successful were you in accomplishing what you were asked to do?"). In table 3.1, we can observe that participants felt they performed better under the IO_0 condition ($p < .001$).

	IO_0	IO_1
How mentally demanding was the task?	50 (36.25)	70 (33.75)**
How hurried/rushed was the pace of the task?	25 (36.25)	90 (31.25)***
How successful were you in accomplishing what you were asked to do?	70 (21.25)***	50 (30)
How hard did you have to work to accomplish your level of performance?	50 (31.25)	75 (31.25)***
How insecure were you?	25 (60)	45 (41.25)*
How discouraged were you?	15 (21.25)	30 (56.25)***
How irritated were you?	20 (28.75)	55 (50)***
How stressed were you?	32.5 (57.5)	57.5 (52.5)**
How annoyed were you?	20 (26.25)	57.5 (55)***
Overall Raw NASA TLX	34.44 (29.3)	59.72 (28.75)***

Table 3.1: Subjective measure - Raw NASA-TLX scores. '*' indicates p < .05 , '**' indicates p < .01, '***' indicates p < .001.

The results show that during the IO_1 condition, the participants felt a higher cognitive workload than during the IO_0 condition (Figure 3.14). The median score for the Raw NASA-TLX in the IO_1 condition is 59.72%, while under the IO_0 condition is 34.44%.

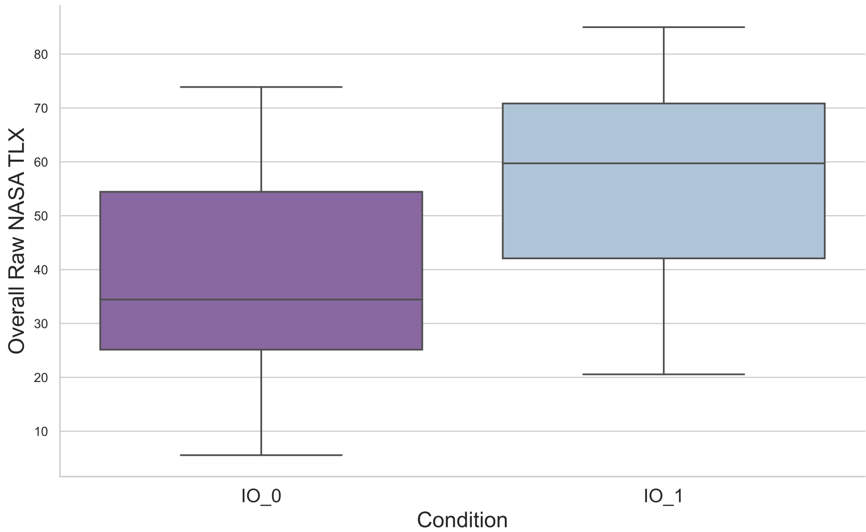


Figure 3.14: Comparison of Raw NASA TLX total scores across IO_0 and IO_1

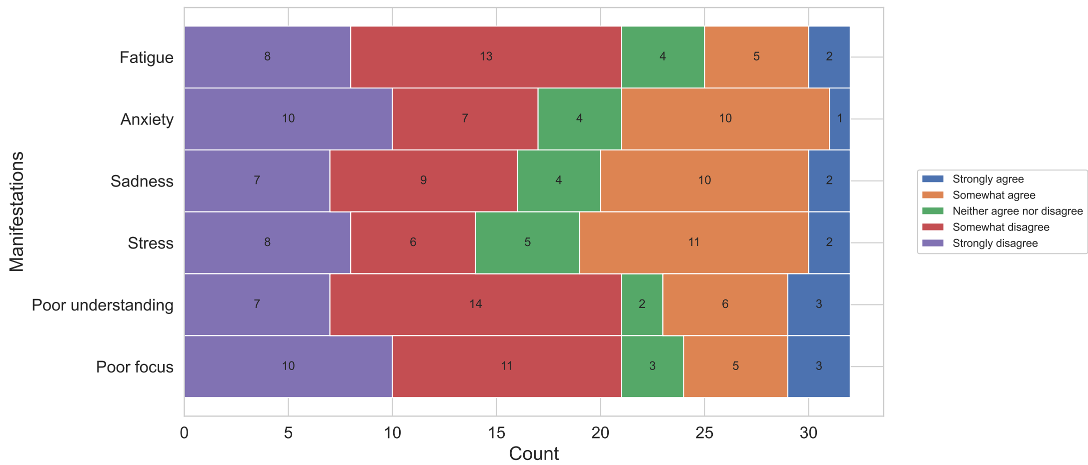
3.4.4 Manifestations

In this section, we look at the manifestations which were identified as relevant to our task and to be perceived by our participants. Across the manifestations identified [9], we only keep: fatigue, anxiety, sadness, stress, poor understanding, and poor focus.

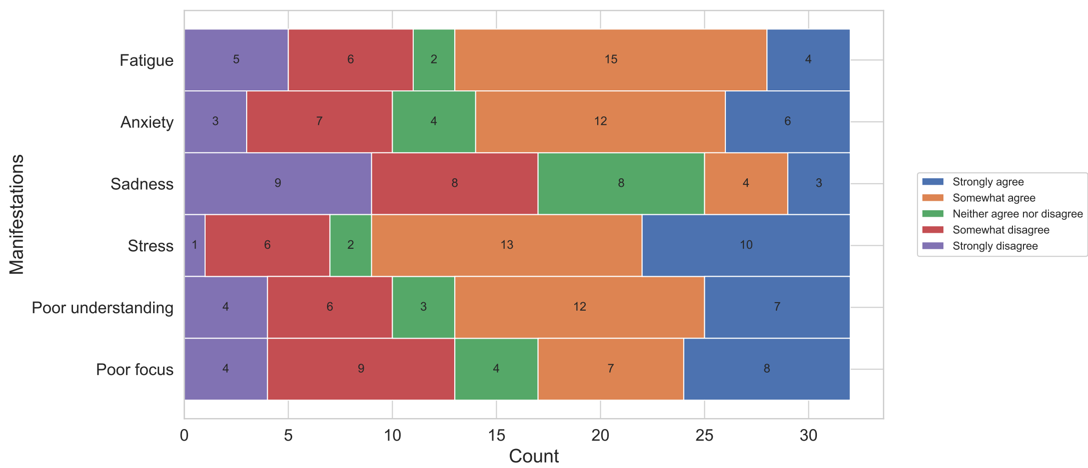
Manifestations	IO_0	IO_1
Fatigue	2 (1.25)	4 (2)**
Anxiety	2 (3)	4 (2)**
Sadness	2.5 (2)	2 (2)
Stress	3 (2.25)	4 (2)**
Poor understanding	2 (2)	4 (2)*
Poor focus	2 (2.25)	3 (2.25)**

Table 3.2: Subjective measures - Manifestations impact. '*' indicates p < .05 , '**' indicates p < .01.

From the results in Table 3.2, we found that our participants were more prone to fatigue, anxiety, stress, poor understanding, and poor focus under the IO_1 condition. While sadness did not seem to be significantly different across the two conditions (p = .22). It is possible that participants were less prone to feel sad since the events did not directly impact them and therefore, participants might feel some emotional detachment. In addition, the fact that both topics were sad, it is likely that the same amount of sadness was induced for some participants due to the nature of the topics rather than due to IO itself. In Figure 3.15, we can see that actually about as many participants were likely to disagree they felt sad under the IO_0 condition (N=16) than under IO_1 (N=17).



(a) Stacked histogram reporting the participants' manifestations' survey results when under IO_0.



(b) Stacked histogram reporting the participants' manifestations' survey results when under IO_1.

Figure 3.15: Stacked histograms reporting manifestations' survey results across both conditions

3.4.5 Triggers

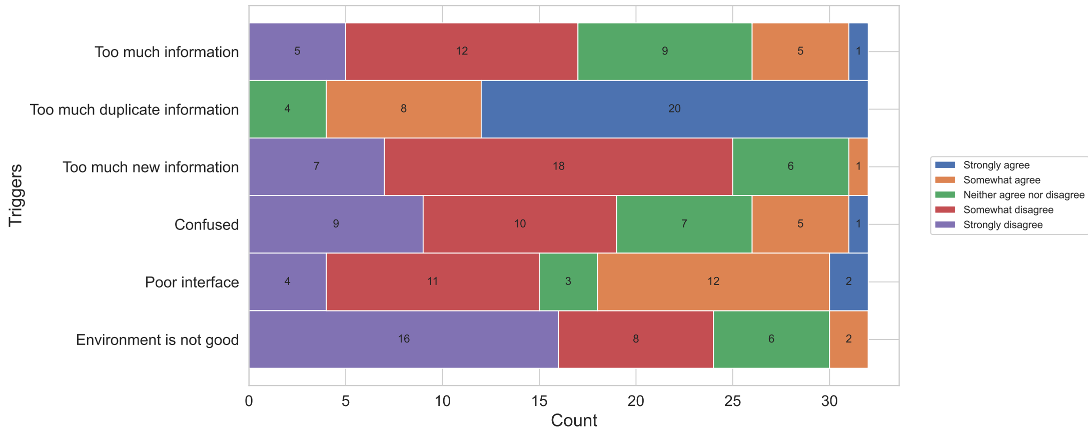
Based on the work presented in the previous chapter, we used the triggers to create IO and the manifestations identified to evaluate whether we were able to create IO or not. In this section, we present the triggers that were relevant to our task and how they were perceived by our participants.

Across the triggers identified, we only kept: Too much information, too much duplicate information, too much new information, confusion, poor interface, and poor environment.

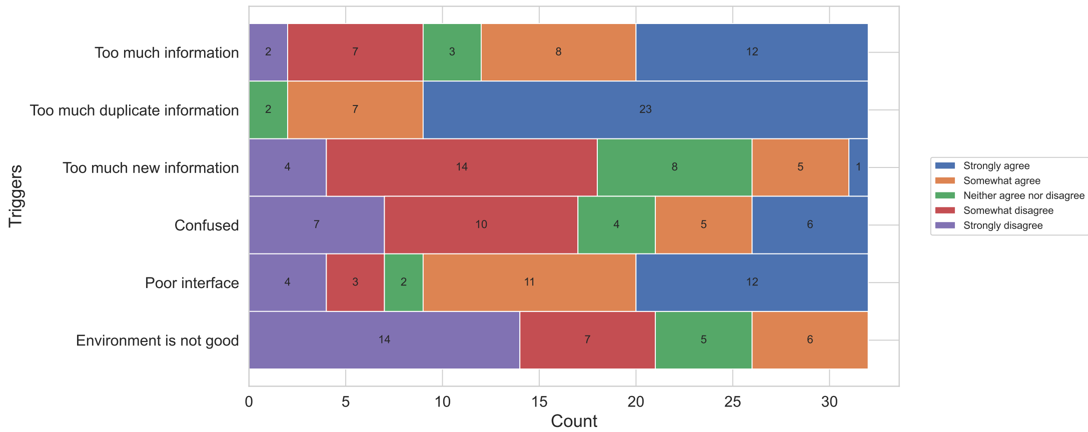
Triggers	IO_0	IO_1
Too much information	2 (1)	4 (3)***
Too much duplicate information	5 (1)	5 (1)
Too much new information	2 (0)	2 (1)*
Confusion	2 (2)	2 (2)*
Poor interface	3 (2)	4 (2)***
Poor environment	1.5 (1.25)	2 (2)*

Table 3.3: Subjective measures - Triggers impact. '*' indicates p < .05 , '**' indicates p < .01, '***' indicates p < .001.

Across all of our triggers, we find that participants were more prone to feel them under IO_1 than IO_0. While for most of our triggers we find a statistically significant difference across the conditions, this is not the case for duplicate information. Duplicate information has a high median ($Mdn = 5$, $IQR = 1$) for IO_0 and IO_1, this indicates that regardless of the condition, participants felt exposed to duplicate information ($p = .3$). Since our events were both selected due to their limited timeframe and presented from the beginning of the incident, it is a safe to assume that participants were going to be exposed to duplicate information and in this case, the tweets would be repeatedly giving the same information regarding the number of casualties and details of the incident. In table 3.3, we observe that across three triggers (too much duplicate information, too much new information, confusion), we have an equal median: Yet we still reach statistical significance in 'too much new information' and 'confusion', this is due to the fact that the Wilcoxon test is a rank sum test and not median test [192].



(a) Stacked histogram reporting the participants triggers' survey results when under IO_0.



(b) Stacked histogram reporting the participants triggers' survey results when under IO_1.

Figure 3.16: Stacked histograms reporting triggers' survey results across both conditions

The stacked histograms in Figure 3.16 are a good way to visualise and understand why we have statistical significance when looking at the 'too much new information' and 'confusion' triggers. We find that more participants were inclined to disagree that they felt those triggers impacted them more during IO_0 (N=25 for 'too much new information', N=19 for 'confusion') than during IO_1 (N=18 for 'too much new information', N=17 for 'confusion'), while more people agreed these 2 triggers impacted them under IO_1 (N=6 for 'too much new information', N=13 for 'confusion') than under IO_0 (N=1 for 'too much new information', N=6 for 'confusion').

3.4.6 Topic preference

Based on the findings above, we look at how the IO condition could impact topic preference. Another potential indicator we have to identify whether we successfully were able to achieve IO is the topic that was preferred at the end of the experiment. At the end of each experiment, we asked the participants to tell us which topic was easier to narrate and why.

We found that the majority of the participants chose the topic which they performed under IO_0 condition, 14 for the Paris attack and 15 for the Boston marathon. While under IO_1 condition, 1 participant chose the Paris attack and 2 participants chose the Boston marathon attack. When we look at the participant which chose Paris as a preferred topic, the participant picked the Paris attack and they attributed the choice to familiarity with the topic:

"I was more familiar with the Paris attacks"

Similarly, we find a similar justification from one participant who chose the Boston marathon attack:

"I'm very familiar with the incident."

The other participant preferred the Boston marathon attack due to the language of the tweets. Most tweets were in English for that topic:

"it was only in English"

3.4.7 Behavioural data

In this section, we look at the behavioural data generated from the participants' interaction with the system. We investigate the engagement with the tweets displayed and the mouse movement changes across the two conditions. For both IO_1 and IO_0 data, we remove the first 135 seconds since it is our ramp-up period for the IO_1 condition.

Likes and tweet hovering

To evaluate users' engagement with the tweet displayed, we have two outcome measures: Likes count and tweets hovered on (Table 3.4). For both measures, we normalise the values by the number of tweets displayed in each condition: Since more tweets were displayed in the IO_1 compared to the IO_0.

Across the two conditions, we find that during the IO_0 condition, participants had more likes ($p < .0001$), they hovered on more tweets ($p < .0001$), and hovered on more duplicate tweets ($p < .0001$).

	IO_0	IO_1
Proportion of likes	.05 (0.03)***	.0009 (0.003)
Proportion of tweets hovered on	.78 (0.35)***	.27 (0.14)
Proportion of tweets hovered on more than once	.29 (0.25)***	.03 (0.04)

Table 3.4: Interaction with tweets. '*' indicates p < .05 , '**' indicates p < .01, '***' indicates p < .001.

Mouse tracking

	IO_0			IO_1		
	Overall	First	Second	Overall	First	Second
Time spent on feed	454.59 (23.58)	457.89 (33.98)	453.86 (23.12)	458.05 (11.21)	458.15 (11.96)	458.05 (11.1)
Travel distance	17914194.58 (15774751.23)*	18992660.57 (15162648.91)*	15250092.1 (15774751.23)	13804364.71 (15194711.41)	14183285.2 (15547398.94)	12428591.88 (15515336.44)
Regions explored	81.0 (36.0)***	81.0 (32.25)**	86.0 (46.25)*	58.5 (33.0)	70.0 (28.25)	52.5 (36.5)
Mean fixation time	0.55 (0.11)	0.57 (0.1)	0.53 (0.1)	0.58 (0.11)*	0.57 (0.06)**	0.61 (0.15)
Express fixations count	76.5 (107.0)*	73.0 (109.25)*	76.5 (70.25)	49.5 (53.75)	60.5 (91.0)	44.0 (31.25)
Modal fixations count	57.0 (59.75)	60.5 (60.0)	52.0 (52.25)	40.0 (53.25)	49.5 (57.25)	33.0 (27.25)
Cognitive fixations count	59.5 (46.75)	57.5 (46.0)	60.5 (38.0)	52.5 (55.0)	58.5 (53.5)	46.5 (55.0)

Table 3.5: Mouse movement analysis across IO_0 and IO_1. '*' indicates p < .05 , '**' indicates p < .01, '***' indicates p < .001.

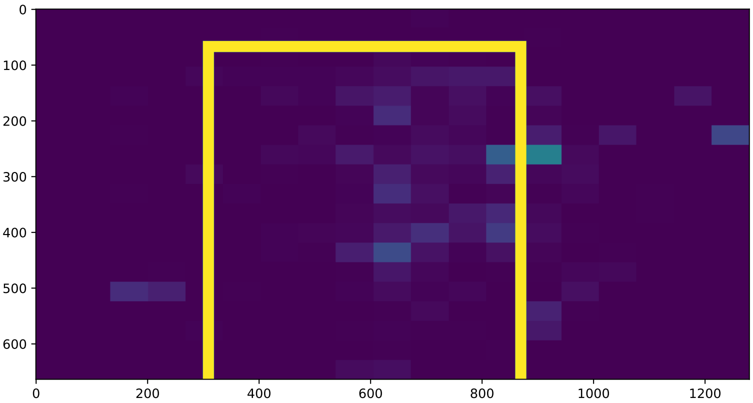
One dimension of our mouse movement analysis consists of looking at the mouse position on the feed during the IO_1 condition and the IO_0 condition. We calculate the total time spent on the feed for each participant across each condition, we find no difference in the total time spent in the timeline area across IO_0 ($Mdn = 454.59$,$IQR = 23.58$) and IO_1 ($Mdn = 458.05$, $IQR = 11.21$), p = .1.

We also look at the total travel distance, we find that our participants moved the mouse across a larger area under the IO_0 ($Mdn = 17914194.58$, $IQR = 15774751.23$) compared to IO_1 ($Mdn = 13804364.71$, $IQR = 15194711.41$), $p = .02$.

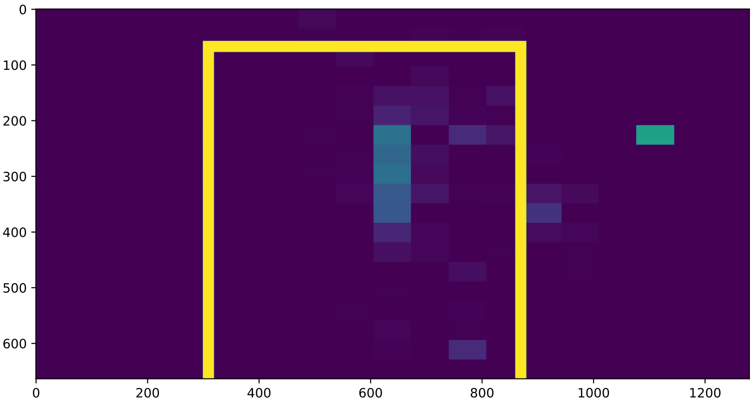
In Table 3.5, we observe that when participants performed IO_0 first, they had a higher travel distance compared to when they did IO_1. However, it is not the case when the participants performed IO_1 first, p = .53. This could indicate that when IO_1 was first, participants were subsequently tired when they had to perform IO_0.

In addition to the time spent in the feed area, and the total travel distance of the mouse cursor, we look at the regions of the screen which are explored (Figure 3.17). The number of screen regions explored under IO_0 (Mdn = 81, IQR = 36) was higher than under IO_1 (Mdn = 58.5, IQR = 33), $p < .0001$. Similar behaviour was found regardless of the order.

Overall, the results for the total travel distance complemented the findings from the total travel distance. In total, 253 screen regions were explored under IO_0 and 188 under IO_1: Participants travelled more with the mouse across the screen and at the same time explored more regions when not overloaded. In Figure 3.17, we visualise the difference in the regions explored. In yellow, we represent the space taken up by the timeline. Within the yellow area we observe that participants under IO_0 moved their mouse all around the timeline area while under IO_1 participants had a more stationary behaviour.



(a) Mouse movement heatmap of the regions explored during IO_0.



(b) Mouse movement heatmap of the regions explored during IO_1.

Figure 3.17: Mouse movement heatmap across IO_0 and IO_1.

While participants explored more regions in the IO_0 condition, they spent more time on a particular relevant region in the IO_1 condition (Figure 3.18), p = .03. This would indicate that after some point, participants were not engaging with the interface and adopted a more passive approach for timeline monitoring. It correlates with the like count which drops significantly in the IO_1 condition.

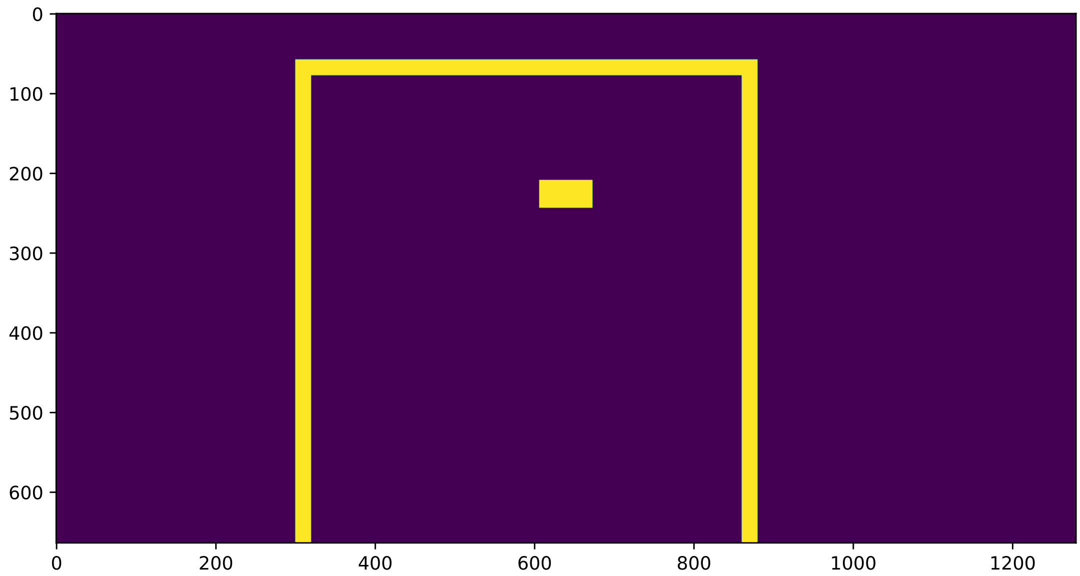


Figure 3.18: Significant screen region under IO_1.

Finally, the last mouse related event we investigate is the fixations. Overall, the mean fixation time was higher under IO_1 ($Mdn = 0.58$, $IQR = 0.11$) compared to IO_0 ($M = 0.55$, $SD = 0.11$), $p = .03$. We find the same trend for when IO_1 is performed first followed by IO_0. However, we do not find a difference when IO_0 is performed first and IO_1 second, p = .56. This could be due to the fact that participants who did IO_0 had more experience with the system and used the same strategy they had under IO_0. In Table 3.5, we report on the three types of fixations we defined (subsection 3.3.14). We find that only express fixations showed a significant difference, participants had more express fixation under IO_0 compared to IO_1. Express fixations are the shortest type of fixations, these results complement the findings from the mean fixation time. Participants under IO_0 had shorter fixations which might indicate that they were less cognitively burdened.

3.4.8 Phases of information overload

Based on Figure 3.10, we have two observations regarding the IO_1: The IO reporting is not evenly distributed across the session, we can see a drop after second 200 and a flattening around 400 seconds. These observations could indicate that IO happens in two phases for our participants. However, under IO_0, the IO reporting is evenly distributed.

In order to confirm that IO happens in two phases, we decide to perform univariate clustering using the K-means method on the IO reporting data in IO_1. In order to determine our optimal k, we used the silhouette evaluation method: It allows us to run our K-means algorithm across a range of values of k and find the optimal k value, with k representing how many clusters we should have.

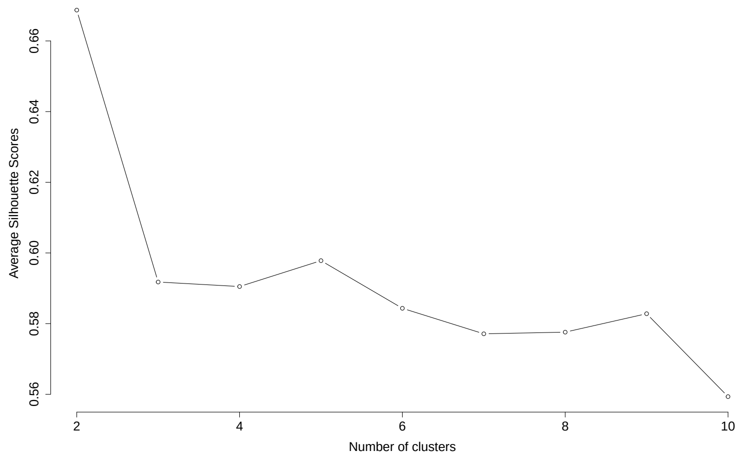


Figure 3.19: Silhouette score of k between 2 and 10

In Figure 3.19, we represent the silhouette score for k between 2 to 10. When k = 2 we have the highest silhouette score, this means that we have two clusters. Once we determined our k, we had to determine the boundaries of our clusters, and based on visual inspection we found that the first cluster started at 0 and ends around 300 seconds, while the second clusters started from 300 seconds to the end of the session (Figure 3.20).

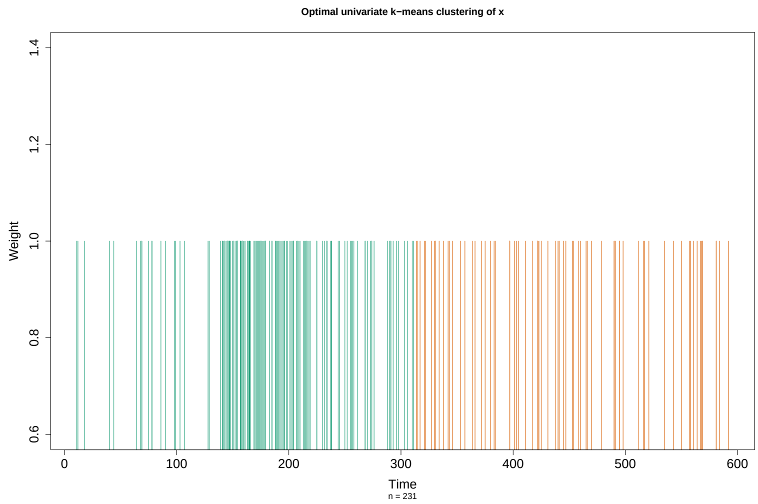


Figure 3.20: The two clusters of the IO reporting during IO_1

The outcome measures related to the tweets (likes and hover on tweets, Table 3.6) show that across phases 1 and 2, participants who were under IO_1 changed their behaviour when interacting with the tweets from the early stages of IO.

Phase 1		
	IO_0	IO_1
Proportion of likes	0.02 (0.01)***	0.0007 (0.001)
Proportion of tweets hovered on	0.32 (0.15)***	0.12 (0.06)
Proportion of tweets hovered on more than once	0.13 (0.11)***	0.01 (0.02)
Phase 2		
Proportion of likes	0.03 (0.02)***	0.0007 (0.002)
Proportion of tweets hovered on	0.45 (0.2)***	0.16 (0.07)
Proportion of tweets hovered on more than once	0.16 (0.15)***	0.02 (0.02)

Table 3.6: Interaction with tweets. '*' indicates $p < .05$, '**' indicates $p < .01$, '***' indicates $p < .001$.

In Table 3.7, we present the analysis across the mouse movement data for IO_0 and IO_1 during the two identified phases of IO. It appears than in the first phase, there are fewer significant differences between the outcome measures. However, once the second phase starts, we find that participants have different behaviour when interacting with the system during overload. In addition to the results present across the two tables, we also find that in phase 1, there are no significant regions between IO_0 and IO_1 while in phase 2, we find two regions where participants spent more time on.

A similar observation to the previous section: It appears that whenever the IO_1 comes first, participants have fewer changes in behaviour and this could be explained by a potential carry over effect: Participants might still feel the overload from the last task.

	IO_0			IO_1		
	Overall	First	Second	Overall	First	Second
Phase 1						
Time spent on feed	161.51 (10.12)	161.51 (8.11)	161.34 (29.65)	161.85 (12.97)	160.51 (9.41)	162.77 (13.39)
Travel distance	4035919.83 (3588069.72)	4197497.12 (2889072.29)	3589825.99 (3595094.82)	4460938.36 (4455669.54)	5209111.49 (3876098.81)	3719790.34 (4439863.06)
Regions explored	53.5 (35.25)*	52.0 (21.5)**	56.5 (38.75)	45.0 (34.25)	55.0 (28.75)	38.0 (23.25)
Mean fixation time	0.53 (0.1)	0.55 (0.12)	0.51 (0.13)	0.54 (0.15)	0.54 (0.12)	0.58 (0.17)
Express fixations count	33.0 (38.5)	31.5 (47.5)	33.0 (29.75)	26.5 (33.0)	35.5 (35.25)	19.5 (27.25)
Modal fixations count	21.5 (19.5)	22.0 (19.5)	21.0 (16.5)*	23.0 (33.5)	37.5 (33.75)	18.0 (19.5)
Cognitive fixations count	23.0 (26.25)	24.0 (24.0)	18.5 (23.25)	23.0 (34.75)	24.0 (25.0)	16.0 (29.0)
Phase 2						
Time spent on feed	289.89 (42.86)*	288.89 (53.9)	289.89 (21.27)	295.35 (10.93)	295.8 (10.21)	294.04 (10.51)
Travel distance	9127356.09 (8379135.96)**	9912754.86 (8265276.38)*	8276894.71 (7282366.09)	5758041.85 (6078707.62)	5869577.77 (7465935.62)	5739452.53 (5100444.66)
Regions explored	72.5 (44.5)***	72.5 (37.5)***	70.5 (50.0)***	47.5 (32.0)	50.0 (28.25)	34.5 (27.25)
Mean fixation time	0.56 (0.1)	0.58 (0.09)	0.52 (0.11)	0.61 (0.09)**	0.6 (0.07)**	0.63 (0.13)
Express fixations count	41.5 (54.0)**	42.0 (54.25)*	40.0 (54.5)*	27.5 (30.75)	29.5 (36.75)	26.0 (29.0)
Modal fixations count	30.5 (37.75)**	32.0 (39.5)	30.0 (29.25)*	19.0 (17.25)	24.0 (20.75)	18.5 (19.5)
Cognitive fixations count	27.5 (24.75)	32.5 (23.25)	25.5 (26.25)	32.0 (28.75)	36.0 (29.75)	27.0 (28.5)

Table 3.7: Mouse movement analysis across IO_0 and IO_1 during phase 1 and 2. '*' indicates $p < .05$, '**' indicates $p < .01$, '***' indicates $p < .001$.

When we compare the findings of the mouse movement with the ones of the behaviour with the tweets, while when interacting with the tweets (like and hover) there is no change between phase 1 and 2, but we do find such change when observing the mouse movement. This finding might indicate that the mouse is a better representation of the users' state.

3.4.9 Session classification

In this section, we use the mouse related outcome measures from the previous section to build a classifier which can predict whether a participant was under IO or not. We built three classifiers, one across the whole task, and two across the two identified phases of IO.

The following features were selected: Number of fixations, average fixation length (fixations are ones greater than 150 ms and less than or equal to 2 seconds), number of express fixations (fixations greater than 150 ms and less than 300 ms), total travel distance, time spent in significant regions, number of regions explored. We evaluated our data on six models: Logistic regression, SVM, XGBoost, decision tree, random forest, and the final model consists of an ensemble model where we combined logistic regression, support vector machines.

We use cross-fold validation as we have a dataset with 64 rows (32 for IO_1), it is a method commonly used for datasets in machine learning: In our case, we use 7-fold cross-validation. We test various number of folds and 7 is the one that had the lowest standard deviation and yielded good results. In addition, to avoid any model bias, prior to feeding our data to our models, we shuffle the values.

Model	**Accuracy**	**F1 score**
Logistic Regression	0.70 (0.1)	0.69 (0.12)
SVM model	0.69 (0.04)	0.68 (0.08)
Decision Tree	0.56 (0.14)	0.54 (0.14)
Random Forest	0.63 (0.15)	0.61 (0.15)
XGBoost	0.56 (0.09)	0.51 (0.15)
Logistic Regression + SVM	0.69 (0.07)	0.67 (0.1)

Table 3.8: Session classification models: Standard deviation is reported between parenthesis.

Across the six individual models we have tested (Table 3.8), we find that the simpler models, logistic regression and SVM, are the ones which perform best on our full data sample (i.e starting from 135 ms). The SVM manages to achieve a fairly small standard deviation which indicates that there are not too many differences across our predictions in each fold. Since we do not have a large number of training samples, going for a simpler model such as logistic regression or SVM is better as it helps avoid the risk of overfitting which is common with neural networks on small datasets.

Subsequently, we use the same features and evaluate the same models but on the two identified phases of IO in our data. We observe in Table 3.9 that our models perform better during phase 2. This is something predictable as we have seen in the analysis across the mouse movement outcome measures that participants had fewer differences in phase 1 between IO_0 and IO_1. Our models perform better across all models and we find that XGBoost has the best accuracy and F1-score. However, we still found that logistic regression, and the SVM model to perform well. Additionally, we found that the decision tree, and random forest outperformed logistic regression and the SVM model during phase 2.

We believe this improvement of XGBoost, the decision tree, and random forest is due to the more significant differences that appear through our data in phase 2. By having more salient difference between the tree nodes, this leads to easier branching out as we go deeper in the tree, decision trees are also known to not be suited where there small variations between the points [181]. Hence, the poor score on phase 1 of the decision trees based models could indicate that a small number of participants had already some effects of IO but it is only in phase 2 that the rest of the participants changed their behaviour as a response to IO. Since random forest is an ensemble model made of trees, the improvement in decision tree is analogue to the improvement on random forest. The same applies for XGBoost which is an ensemble model based on decision trees.

	Phase 1		Phase 2	
Model	**Accuracy**	**F1 score**	**Accuracy**	**F1 score**
Logistic Regression	0.56 (0.17)	0.53 (0.17)	0.72 (0.12)	0.72 (0.12)
SVM model	0.56 (0.22)	0.55 (0.20)	0.72 (0.12)	0.70 (0.11)
Decision Tree	0.56 (0.09)	0.61 (0.11)	0.75 (0.07)	0.75 (0.1)
Random Forest	0.56 (0.1)	0.55 (0.1)	0.77 (0.03)	0.75 (0.06)
XGBoost	0.57 (0.16)	0.54 (0.18)	0.80 (0.14)	0.80 (0.15)
Logistic Regression + SVM	0.56 (0.17)	0.53 (0.17)	0.72 (0.12)	0.72 (0.12)

Table 3.9: Session classification models: Standard deviation is reported between parenthesis.

3.5 Semi-structured Interviews

Our 32 participants took part in a semi-structured interviews at the end of the study. These interviews allowed us to have a better idea of how IO was perceived, the dimensions of IO that were manifested, and how participants tried to adapt by adopting strategies.

3.5.1 Findings

Our overall Cohen's Kappa score reached 0.81 which indicates an almost perfect inter-coder agreement [193]. We achieved a 0.74 Kappa score (substantial agreement) for the emotional manifestations of IO, 0.43 (moderate agreement) for the failure to filter code, and 0.86 (almost perfect agreement) for the adoption of strategies by our participants.

Across all our participants (N=32), we find that they all adopted a set of strategies, one popular strategy is filtering tweets however this was done using different approaches: Filter information based on keyword (hashtag or specific words such as "prayer"), length of tweets, or text format (capital or lower case, presence of emoji).

In Table 3.10, we report some direct quotes related to each of our three codes, i.e emotional manifestations, failure to filter, and strategies.

Emotional manifestations
"Stressed because I thought I might miss stuff"
"I was more irritated"
"I was annoyed and irritated"
"Stress and kind of anxious"
Failure to filter (consequence)
"Couldn't think about fake news, I would look at repeated information"
"Couldn't filter anything in event 1, just went through everything and it was too much stress"
"Sometimes capture all with no filtering as it's more difficult to look at sources"
"Difficult to assess what's good and credible, if it goes so quick"
Strategies
"Checked @ for source, checked for 'rumoured' vs 'confirmed', made sure to not focus on repeating information, didn't scroll because information will repeat"
"ignore tweets with hashtags, look for numbers sometimes"
"go with breaking news keyword, and otherwise ignore hashtags"
"Pay attention to keywords, skimmed over long tweets, read caps fast because emotional response, pay attention to the ones with numbers"
"Look for media outlets and keywords, look at the length of the tweet, stay at the top"

Table 3.10: Quotes from semi-structured interviews across our three codes: Emotional manifestations, failure to filter, and strategies.

3.6 Discussion & conclusion

The aim of this work was to study IO 'in the lab', in order to achieve this we relied on our findings from chapter 2. In chapter 2, we defined the concept of IO and established that extrinsic information characteristics were trigger of IO. In particular quantity and speed: "IO manifests itself through emotional and cognitive challenges and is most likely to happen through intrinsic and extrinsic information characteristics" [9]. Hence, we used frequency of tweet updates as our trigger. We created and demonstrated that IO had occurred by using participants' reporting of the feeling of IO, using our modified NASA-TLX (NASA-TLX that is a commonly used method in IR to measure cognitive load), and semi-structured interviews that have shown that our participants suffered emotional manifestations and consequences of IO as identified in our first study [9]. We found a significant increase of IO reporting as well as in the modified NASA-TLX scores when participants were performing a task under IO condition. In addition, we evaluated how the feeling of IO impacted the manifestations which were identified in our first study [9], we found differences in the ratings across participants with higher scores for IO condition. Our semi-structured interviews allowed us to complement these findings but also highlight the adoption of strategies by our participants. This latter finding should be further studied and a classification of these strategies as either deliberate or emergent, as per the conceptual analysis by Savolainen [74], would allow to see how cognitively individuals differ when reacting to IO.

To answer RQ 1 and RQ 2, we looked at the behaviour of our participants under IO and not IO. An analysis of participants' behavioural changes based on their interaction with tweets (likes and hover) found that they liked less tweets, hovered on less tweets under IO, and hovered less on older tweets under the IO condition. This withdrawal behaviour has been identified in our work [9] as a common consequence of IO: In previous work with Twitter it was shown that in a queue structure first-in-first out, older tweets were lost as new tweets came in as it required a higher cognitive load to seek older tweets [138]. The analysis on the mouse movement data has allowed us to extract differences between the participants who were under IO and the ones who were not. Under IO, time on the feed, travel distance, regions explored, number of fixations (express and modal) all saw a decrease while the mean fixation time saw an increase. Participants also spent more time on specific regions under IO. These behavioural changes based on the mouse movement offer an unobtrusive way to detect whether participants were under IO or not. The increase in mean fixation time under IO could indicate that participants' information collection required more efforts: Visual collection of information requires eye fixations, past work has shown that individuals often use the mouse to guide their attention, henceforth an increase in the mouse mean fixation time would indicate that participants had to increase their efforts to read and encode the information they were receiving. In addition, based on the assumption that the mouse can be used as a proxy for the eyes, the decrease in express and modal fixation with the increase of the mean fixation time matches the findings of previous eye tracking studies [121]. In order to answer RQ 3, we used unsupervised learning on the IO self-reporting during the IO condition. A clustering analysis of the IO reporting data of the participants allowed us to identify two phases, with a first phase lasting from the start of the IO condition (135 seconds) to second 300, and a second phase from second 300 to the end of the study. Based on this finding, we looked at our behavioural data across phase 1 and 2. While the interaction with tweets (likes and behaviour) showed the same changes happening as on the overall session and no difference across the IO condition in the two phases, the analysis on the mouse movement data showed more important changes across the two phases for the IO condition, in addition to the changes across conditions. While our system entered the IO condition at second 135, the change of behaviour from the not IO condition exacerbates during the second phase (second 300): This indicates that the onset of IO at an individual level is gradual. This gives a first insight into the dynamic nature of IO and how it is felt over time. Our last RQ 4 lead us to investigate the features to be used to develop classifiers for IO, hence we focused on mouse tracking data. The mouse related outcome measures allowed us to develop classifiers which leveraged these outcome measures as features. Our classifiers yielded overall better performances in predicting IO during the second phase. We believe that these results could overall be improved with more data. The overall session classification was able to reach 70% with logistic regression, while the first phase classification reached 56% which indicates that the classifiers were not able to discern important differences between the features from the IO condition and not IO condition when early in the IO condition. The second phase classification reached a score of 80% with XGBoost. The important difference in scores between phase 1 and phase 2 confirms what was observed in our analysis of the mouse related outcome measures: While participants noticed the change by reporting IO at the start of the condition, IO manifestations and consequences started to only affect them a few minutes later.

In our review of the previous work with IO looking at the interactions of the users through eye tracking and mouse tracking, we did not find work which actually realised an empirical study involving a real world stream of information which anyone can face every day when using an IR system. Rather, we found artificially designed tasks or tasks borrowed from the fields of neuroscience and psychology (n-back tasks) to study how the behaviour of participants changes under high cognitive load [126, 168, 169]. While these studies are promising, we need more work looking at the same behavioural changes in realistic environments to confirm their validity and whether they translate well.

In this chapter, we have used the findings from our previous study to design a user evaluation which can show how to create IO in the context of a popular social network, Twitter. We use quantity of information and speed of information as our triggers. These triggers are extremely common in the context of social networks especially on Twitter. Due to the design of the platform, users will actively comment on news but also share details about any on-going event they are part of. It becomes a source for journalists to actively collect information but also for first-aid responders to draw out a clear picture of the state of an event and locate any potential casualty. These two examples are prone to use tools such as TweetDeck which allows a live update of the timeline. However, platforms such as TweetDeck will put users under an IO state and this might affect the ability of the individuals to filter out the information they receive and might have consequences in the spread of misinformation or in more important cases, on the lives of the responders and casualties. In addition, we have used identified manifestations from our concept analysis to determine whether IO happened. These manifestations complement the results from modified NASA-TLX, we believe these can be used in the future to assess whether users of an information system were under IO and design a more CIS tailored questionnaire to measure IO. The explicit and implicit behavioural data to show the changes that happens when an individual is overloaded does not require additional equipment and offers an unobtrusive and scalable way to detect when individuals are under IO. While some SNS are offering ways to set time limits to support individuals' well-being, such manually controlled features can add a cognitive burden to the user and often results cannot be used. Using our approach can offer a way to design intelligent, adaptive systems which can adjust to each and every individual.

3.6.1 Limitations

While we believe this is a pioneering study in the field of CIS with the intention of understanding IO, it is only a small step into the disambiguation of the concept. We limit the presented tweets to text format only, we remove images and videos as this would be an additional modality which might have a different impact on how IO is ignited within the participants. Often, when interacting with a system like Twitter, individuals would click on links to find out more about a topic. However, we did not allow clicks as we wanted to ensure the scope of our study did not use additional resources which might affect the IO state. Finally, while we use the mouse as a guide to study how the user behaves and where they might be looking, it has been shown that the mouse can be stationary while the user looks somewhere else [165], however in our case we keep the user engaged with the like demand, so the stationary state an individual might enter was limited.

An additional limitation of this study is related to the participants' education level. Since we used convenience sampling in recruiting our participants, the majority were students and hence do not necessarily represent the general population in the way they use social networks.

This study was conducted in 2021, due to the COVID pandemic we had to set up our study online as we did not have access to on-campus facilities. We used Zoom (an online video conferencing tool) as our interface to allow the participants' to control the screen where Twitter was running. While we tried to limit any latency, there might have been a lag that is due to the way Zoom renders the study lead's screen which would have not occurred if the study was all happening locally i.e without the videoconferencing tool. This latency would be linked to the connection of the host computer but as well as of the participants. However, since our data was written locally, it can be safe to assume there was no additional latency at the host machine level. In addition, we did not observe any important variation in the behavioural data across our participants which would indicate that if latency issue occurred due to Zoom, it was minor.

3.6.2 Summary

In this chapter, we tested the findings from chapter 2, and were able to use extrinsic information characteristics as our trigger for IO. The resulting IO state of our participants allowed to extract behavioural changes due to IO. A clustering analysis allowed us to identify that IO perceived is differently over time and that its effect are not instantaneous. Finally, we leveraged outcome measures related to the behavioural changes of our participants to classify when individuals were under IO.

In the next chapter, we approach the problem of IO from a different perspective. We borrow from research in cognitive neuroscience, and attempt to tackle IO using neuromodulation.

Chapter 4

Neuromodulation: A novel approach to tackling information overload

In this chapter, we move away from the system-side view of tackling information overload (IO) to a user-based view of the problem. Hence, we started by reviewing work both in information retrieval (IR) and cognitive neuroscience, and identified brain regions involved in the search process which are relevant (section 4.1). We used neuromodulation to stimulate one specific brain region which is involved in executive functions (in this case working memory) and has been identified as active during the search process in prior IR studies: The left dorsolateral prefrontal cortex (DLPFC). We designed a clone of Google (Moogle), and we aimed to use a neuromodulation device, transcranial direct current stimulation (tDCS), to reduce the dwell time which is the time spent on web pages (excluding the search engine results page), the dwell time which is associated with relevance judgment (section 4.2). By stimulating the left DLPFC, we believed the improvement in working memory would result in supporting our participants' information load and by remembering more information throughout the search process, they would be able to filter out through information faster and better direct their attention. However, we did not find any significant change in the dwell time and a further exploration of additional outcome measures derived from our log data did not show any change across any of the additional outcome measures (section 4.3). We did find that the side effects associated with tDCS did moderate tDCS effect which would indicate that an online stimulation design could affect the results of studies using tDCS. We offer a discussion at the end of the chapter where we present potential limitations as well as recommendations for future studies using tDCS (section 4.4).

4.1 Introduction

Traditional approaches to tackling IO and improving the online search experience have been through system-side improvements, whether through the interface or the algorithms powering information retrieval systems [194]. In Computer and Information science, these improvements have been identified through various sub-fields such as IR and human-computer interaction [195]. One of the most notable improvements in recent history in IR was through the algorithm PageRank [6] which introduced a better way to rank web pages and reduce noise for users. Most recently, with the development of neural networks, we have seen a shift from traditional algorithms to new machine learning powered systems which can offer more relevant content to the user by combining latent features of the users and different signals (queries submitted, page visited, geographical localisation etc) [33]. In this context, one of the most important transitions from traditional algorithms to neural networks was when Google started using Transformers to improve searches on its search engine [196].

In this chapter, we shift from these system-side improvements and look at how we can leverage advances in cognitive neuroscience to tackle IO from the users' side. In our attempt, we focus on one type of search task in IR and interactive information retrieval (IIR) which is exploratory search.

4.1.1 Information Retrieval

Historically, humans sought information either through direct interactions or through written communication. While cataloging schemes have been around since the third century [197], it is in the 1940s the retrieval and storage of information became a pressing matter for the scientific community [19]. The need for such systems was first articulated by Vannevar Bush in 1945 [19] with the description of what he called the "memex": "A memex is a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory."

With the development of the internet, there has been an explosion of search systems that enable fast retrieval of information anywhere and at any time. Without a doubt, the most popular search system is Google[1].

The subfield of Computer & Information Science which looks into the development of these systems is IR. The term 'information retrieval' was first used in 1950 by Calvin Mooers [198] to describe the process of looking for information: "The requirements of information retrieval, of finding information whose location or very existence is a priori unknown..."

Today, we can describe IR as the field that looks into the representation, storage, indexing, and retrieval of documents [199]. Two major developments in IR were indexing and ranking [5]. Regarding indexing, there was the development of the Uniterm system [200], each document was indexed using a list of keywords. These keywords were then used to retrieve the relevant documents through a Boolean search. A Boolean search consists of building a query using boolean operators (example: OR, AND, NOT...etc), the following query would return all the documents which have the keyword 'Scotland' associated with it but not the keyword 'Glasgow': "Scotland" AND (NOT "Glasgow"). The problem with such a retrieval approach is that there is no ranking associated with the returned results. Another important advance in IR was the introduction of ranking, an approach that required manual weight assignment to keywords is *term frequency*: It assigned a weight to each keyword of a document based on how important each keyword was to the document [201]. Later on, the term frequency-inverse document frequency (tf-idf) was developed which lead to moving away from the manual approach to a method that relies on calculating the number of times a word appears in a document and assigning it a lower weight [27]. The indexing of documents also shifted

[1]Google.com accounts for about 70% of the search engine market share for desktops/laptops and for 94% for mobile devices as reported by https://netmarketshare.com/ (last accessed: 22/09/2022)

from a list of keywords to an approach known as bag-of-words where each document is mapped to an N-dimensional space.

When the World Wide Web was created and with the growing amount of websites, there was a need to move from manual cataloging to indexing the various web pages [5]. One of the most popular approaches to this problem leveraged the links of individual web pages to others (link analysis), it is known as Pagerank [6].

While IR is the field that deals with the system view of search and since it has seen continuous growth, it lead to a new field of research which is IIR. The latter aimed at building a theoretical view of the search process by integrating the searcher's journey and interactions with IR systems. By better understanding the searcher's journey, the IR systems developed can better answer the information need of the searchers.

4.1.2 Interactive Information Retrieval

IIR is the field that studies how individuals use search systems to retrieve information [202]. There are two types of searches, the first one follows the lookup model which is about giving a specific answer to a specific question (i.e: question-answering, fact-finding), and the second type is the exploratory search, which is all the other types of searches [89]. In figure Figure 4.1, we present the search activities as articulated by Marchionini and White [1]:

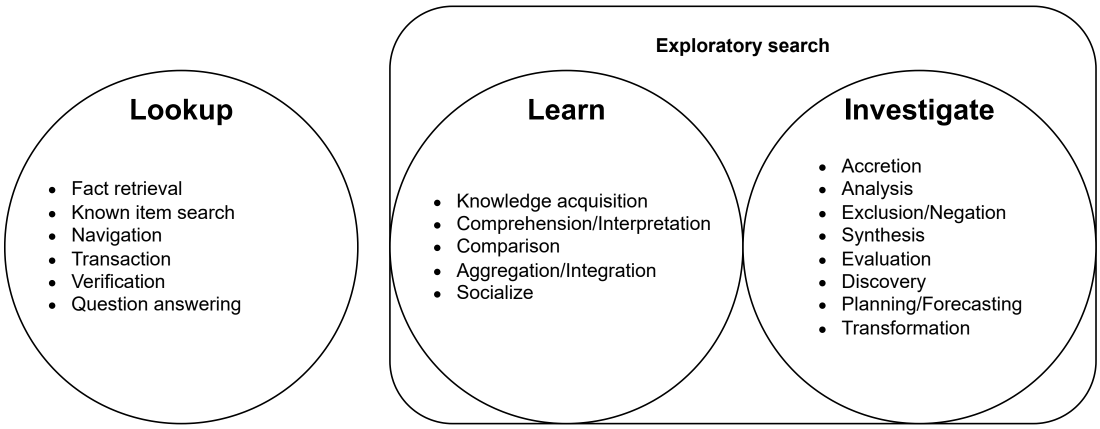


Figure 4.1: Classification of search activities as per Marchionini and White [1]

One example of a lookup search task is trying to find the answer to the question:

'What is the capital of Morocco?'. It has a straightforward answer (Rabat) and does not require additional exploration. While exploratory search is a type of search which has different steps involved as part of the search process. One type of exploratory search task would be having to find out about 'The Fall of Constantinople' as part of a report for school.

Prior to the start of a search, the searcher has to recognise and specify their information need which is then followed by the examination of search results and, depending on the complexity of the information need, the searcher has to repeat this process until their knowledge gap is filled [1, 203].

To better understand the information-seeking process, researchers offered various models that [204, 205, 206, 207]: Belkin [207] specified the Anomalous State of Knowledge (ASK) hypothesis which states that there is a gap between the knowledge of the searcher and what they are looking for, this gap is the driving force of the information-seeking process. In her seminal paper, Bates [204] described the search process as berry-picking where the searcher picks up information continuously and this drives their next step in the search process.

The various concepts associated with the models of the search process describe the use of search strategies and the involvement of a learning dimension: As such, there is a cognitive effort involved that will vary at the various steps of the search process, a cognitive effort which will involve various brain regions. Our work focuses on one component of the search process which is the collection of information from the web pages found during a search task, and how in this case we can support the searcher in one type of search tasks: the exploratory search.

4.1.2.1 Exploratory search

Marchionini [89] distinguishes three search activities. The first one is labelled as lookup while the other two are learn, and investigate. While the lookup is characterised by a clear goal, as the searcher is looking to find an answer to a well-defined question, the learn and investigate search activities are encapsulated under what is called exploratory search. We introduce and present some of the many definitions of exploratory search below, and by characterising this type of search we can highlight and use its specific characteristics to identify later on neural correlates which are likely to be involved.

Exploratory search is an advanced type of search that has been described by Marchionini [89] as any search task that is not lookup, it involves complex cognitive skills and knowledge acquisition [18]. One example of an exploratory search task is when you are trying to perform a literature review on a specific topic, you have a poorly defined information need and you are not sure of what would be the boundaries within which your specific topic falls in. It requires an iterative dimension to search, you refine your queries and identify relevant academic papers which can help you understand better the topic.

White and Roth [18] characterised exploratory search by an ill-defined goal, curiosity, and uncertainty. The fundamental nature of exploratory search requires search engines to provide more support to the searcher to overcome uncertainty [89, 208].

Hendahewa and Shah [209] have shown that query suggestions in exploratory search were not enough to support the searcher and recommended search trails as more effective tools to reach preferable search outcomes. Due to the ill-defined information need the nature of exploratory search, it is difficult for the searchers to determine the relevance of documents.

Recent work by Soufan et al. [210] showed that literature reviews can be described as exploratory search. The authors have shown that there is a fourth dimension to exploratory search which is knowledge gain/change, it adds itself to the user dimension, the problem context dimension, and the search process dimension. The fourth dimension identified drives the iterative process of the exploratory search and can affect its quality as the searchers' knowledge is evolving throughout the search process, they use their newly gained knowledge to determine the relevance of the information at hand. The importance of relevance feedback was already highlighted by Marchionini [89] as key to improving the exploratory search process but its main challenge was that searchers would not provide explicit feedback about the relevance of a result and therefore the search engines of the future would have to mine the searchers' behavioural data to determine if a result is relevant or not.

As we have seen, the search process is an iterative process and relies on the users' interaction with IR systems in order to better support the users' information need. One heavily studied component that helps systems better support the users is relevance.

4.1.2.2 Relevance judgment

In the past, various efforts were made to define relevance judgment (RJ) and collect relevance feedback in attempts to improve the search experience.

The first efforts from the IR community were made by explicitly requesting feedback from the users. In explicit relevance feedback, users were asked to explicitly say whether a document is relevant or not. This would allow us to suggest newly expanded queries [211] and improve recall (i.e. provide them with more relevant corpora [212]). However, explicit relevance feedback placed a cognitive burden on the users, who could also be reluctant to provide explicit feedback. Besides, these approaches could be expensive, as the new queries can be long and therefore take more time to retrieve. As explicit relevance feedback has shown its limitations, there has been a push in research towards determining implicit relevance feedback [213, 214, 215, 216, 217, 218] and/or affective feedback [219]: One signal of implicit feedback would be the use of the dwell time (time spent on web pages excluding the time spent on the search engine home and results' web pages) to determine whether a person finds a web page relevant or not and affective feedback would be relying on facial expressions.

The relevance of the discovered items is correlated with session-level satisfaction [220] and from a user-oriented view of relevance, it is subjective and depends on the individual mental experience, henceforth it will involve a cognitive restructuring where the searcher's perception of the relevance of a specific item can change as they progress in their search [221, 222]. In parallel to this subjective view of relevance, we find a system-oriented view that treats relevance as a static and an objective concept; in this class we find the work by Saracevic [223]. In 1975, Saracevic highlighted that to understand relevance we have to question its nature [223]. Based on the question 'What is the nature of relevance', recent attempts have been made to understand it through physiological approaches [157, 224].

4.1.3 Neuro-Information Science and NeuraSearch

Over the years, various research fields have borrowed from Neuroscience to build better theories around human behaviour [225, 226, 227]. In the case of CIS, we have Neuro-Information Science [146] and NeuraSearch [228]: Two fields that are looking into using neurophysiological devices to complement traditional data collection methods and improve the searchers' experience.

Recent work in Neuro-Information Science focused on studying neuropsychological research tools to better understand the steps involved in the search process and with a further goal to eventually create neuro-adaptive IR systems [146]. A new and growing body of work has looked into better understanding the neural correlates involved in information need, and information relevance. Other work has looked at studying how knowledge gain affects users' eye movements through eye-tracking devices [147].

Work done in identifying the brain regions involved in the search process can be leveraged in the future with the development of brain-computer interfaces [229].

The majority of work done in the intersection of Neuroscience, IR, and IIR has been around the concepts of information need and information relevance. As discussed, both of these concepts have been key to understanding the information-seeking process.

While Neuro-Information Science encompasses the work done with all the various neurophysiological research tools (eye-tracking, electrodermal activity, pupil dilation, electroencephalogram, Functional magnetic resonance imaging), we find that NeuraSearch focuses on the neural correlates involved in the search process [228].

NeuraSearch

The two most commonly used devices to explore the brain in NeuraSearch are: electroencephalogram (EEG) and functional magnetic resonance imaging (fMRI). EEG offers a real-time measurement of brain activity: The electrical activity within the brain represents the rapid electrical impulses that neurons use to communicate with each other. EEG can detect these signals non-invasively on the scalp. The signals are either analysed with respect to the frequency domain, where each frequency is associated with a specific cognitive or motor process, or into event-related potentials (ERP).

ERPs are generated after the display of a stimuli, stronger amplitude of the ERPs means more neurons are active within a region.

EEG offers a high temporal accuracy, and through source localisation, it can offer good spatial accuracy. In the field of NeuraSearch, to achieve higher spatial accuracy, researchers use fMRI with Blood Oxygen Level Dependent (BOLD) imaging which consists of looking at cerebral blood flow to detect activity differences in the various brain regions. Using fMRI offers very high spatial resolution, however, it is an indirect measure of activity since it has to wait for activity in a region and for the blood flow to reach the given region and it is also very slow, due to the slow scanning there is no real-time insight on which brain activity is activated at a given time, another major limitation is that participants must lie down and stay still during the scan which limits our ability to develop and test realistic IIR tasks. While EEG offers, through source localisation, both high temporal resolution and good spatial resolution, it is also cheaper and more flexible in that participants can conduct the study in more natural positions.

One of the first concepts from IIR to be studied extensively using Neuroscience approaches is relevance [38, 224, 229, 230]. Early work combined affective, behavioural and physiological data to predict relevance of search items, in this case the items were videos [38]. The researchers explored new elements, which can complement implicit relevance feedback (dwell time) and replace task information, which is often combined with the dwell time, to determine relevance. The potential complementary signals presented in the paper are affective and psychological signals. The affective signal was facial expression and the physiological signals combined various sensory channels, such as brain signals, for which they used NeuroSky MindKit-EM. This processes EEG signals and returns values to measure attention and meditation levels, additional physiological signals were heart rate data, galvanic skin response, and skin temperature. They found that physiological signals were as good as task information and that the dwell time with physiological signals are reliable to replace task information.

Allegretti et al. [230] aimed to identify the timeframe within which relevance judgement happens within the brain: They used EEG and devised a task where participants had to decide whether an image was relevant to a specific topic. Based on this they compared the brain activity when an image was relevant and when it was not. Their primary aim was better at defining relevance and its implementation. Since relevance is subjective and multi-dimensional, searchers used query words to retrieve information, however those words can be uncertain and noisy. In order to improve relevance, a progressive disambiguation is needed, this process is interactive and iterative. To achieve this progressive disambiguation they looked at two types of feedback systems: Implicit (e.g dwell time) and explicit (e.g asking individuals to mark a search item as relevant or not). Explicit feedback is cognitively difficult to use and for this reason, implicit feedback was employed. Implicit feedback can be noisy but allows to collect distinct signals of the user's intentions (actions and user generated content). For example, in the case of the dwell time, individuals can access a web page but walk away, hence this would result in a high dwell time which would not accurately reflect the actual individual's relevance judgment. To improve the implicit feedback they used EEG brain signals as their dependant variable and relevance as the independent variable in their study.

In their EEG based study, Allegretti et al. [230] used a 64-channel EEG system and relied on the mean significant difference between the relevant and non-relevant conditions to measure the difference in activity: They subtraced from the mean signal of the relevant condition, the mean signal of the non-relevant condition. The EEG brain signals were split into four different epochs: In the first epoch (0 to 180ms), no difference between the two conditions was found, between 180 to 300ms there was a higher mean significant difference under the relevant than the non-relevant, where a stronger amplitude was found under left frontal (F1) and right anterior-frontal (AF4) electrodes. Between 300 and 500ms, there was activity shifting to the right central electrodes (C2 and CP2) with CP2 having the greatest mean significant difference between the two conditions. However, the timeframe 500 to 800ms had the largest mean significant difference around the vertex for relevant items compared to non-relevant items, specifically under C2, Cz and C1. In their study, Allegretti et al. [230] highlighted that they derived the first epoch from the work by Thorpe et al. [231] which identified an ERP happening around 150ms indicating visual processing which could be the N170

ERP. The 180-300ms epoch's ERP is not defined in their study but could be related to the N200, the N200 has been associated with the identification of stimuli and executive control [232]. They associate with 300-500ms epoch with the P300 ERP which is associated with relevance of stimulus [233]. The last epoch, 500-800ms, ERP was not defined but could be the late positive component which is associated with recognition and evaluation of stimulus [234]. Moshfeghi et al. [224] tackled the problem of relevance through using a simple task based on topical image relevance judgement during fMRI. Participants were given a topic description and had to asses whether the image stimuli was relevant to the topic description or not. In the study, they found greater activation for relevant stimuli than non-relevant stimuli in the superior frontal gyrus, inferior parietal lobe and the posterior region of the inferior temporal gyrus near the occipital cortex. The activation in the frontal and parietal regions of the right hemisphere are associated with visuo-spatial working memory, and this matches the task demands, as it was designed with images as the target modality for relevance. More recent work was able to move a step forward and identify a graded-relevance, a move away from the traditional binary encoded concept of relevance (relevant/non-relevant). A study using EEG was able to identify three levels of relevance: high, low, and non-relevant. This study relied on previously identified ERPs that were involved in the concept of relevance: (1) P300 is a positive potential that occurs at least 300ms after the stimulus on-set, with a greater P300 meaning greater relevance of stimulus [233], (2) they used the N400 which has negative potential for relevant than irrelevant stimulus and occurs 400ms after the stimulus on-set, (3) and finally the P600 which had a positive potential for relevant stimulus [235]. They found that the amplitude of the ERP components varied depending on various levels of perceived relevance [236]. In these studies, it was demonstrated that there was a clear difference in neural activity across the scalp when making relevance judgment with a difference in neural activity based on varying levels of relevance.

Another key concept which was studied by the NeuraSearch community is information need [237, 238]. Moshfeghi et al. [237] identified activations in the posterior cingulate during information need. More precisely, the dorsal posterior cingulate region where the activity was greater for the information need condition. The dorsal posterior cingulate region is activated when directing attention towards external sources: In this case, participants of their study have to shift from internal information to external sources as they engage and collect new information.

An important work in NeuraSearch aimed to create a model of searchers behaviour using fMRI by using the steps involved in the search process [124]. In the past, attempts to create such models relied on questionnaires, interviews, interaction of searchers with IR systems, we find important models born from these attempts: Belkin's ASK model [207], Wilson's information-seeking behaviour model [239], and Ingwersen cognitive model [206]. Different theoretical models have divided the search process in different ways. Saracevic defined the stratified model where the interaction of the searcher with the IR system is seen as a dialogue between the searcher and the computer through an interface at a surface level [240]. Process models consider the search process as a multi-stage representation of the activities of a searcher. Finally, the cognitive models choose to divide the search activity into the cognitive processes associated with it. In their paper, Moshfeghi and Pollick [124] divided the search process into the different cognitive components which compose it. The cognitive components in the model they present were based on the brain activity as the searcher proceeded with their search. Through this fMRI study, the researchers explored the dynamic allocation of neural resources to different functions and how there are different brain activities depending on which stage of the search process the searcher is at. The experiment used a within-subject design and the model of the search process was divided into five epochs: (1) Information need, (2) query formulation, (3) query submission, (4) relevance judgment, and (5) satisfaction judgment. The four transitions between each epoch were studied and used as the independent variables. The dependent variable was the brain activity revealed by the BOLD signal. They were able to identify important clusters involved in each step by looking at the changes in brain region activations at the different steps of the search process. The following are the networks identified: Visual system (in charge of visual processing), somatosensory and motor system (action control and touch), dorsal attention and ventral attention network (directs attention), fronto-parietal and cingulo-opercular network (cognitive control), salience (determining value of stimuli), default-mode (when attention is internally directed), and the auditory system (related to hearing, and processing sound).

These findings contributed in mapping brain regions to the different stage of the search process. However, the task designed for this study relied on the lookup type of search tasks (e.g "Who is the president of Stanford university"), which has a definitive answer.

While the aforementioned work focused on identifying brain activations during IR tasks and understanding the role of the various brain regions at the different stages of the IR process, particularly for IN and information relevance, this knowledge needs to be leveraged to help improve the users' interaction and inspire the development of new intelligent adaptive systems [229, 233, 241]. One work that successfully leveraged this knowledge is by Eugster et al. [229]: They used brain signals collected using EEG to develop a predictive filtering model, the brain signals were used to detect whether a word was relevant or irrelevant to a task to recommend a set of documents.

So far, we have seen the various characteristics of the exploratory search task as well as the work done in identifying neural correlates involved in the information-seeking process, the work as highlighted focused on tasks which did not involve exploratory search. Among the important aspects of the exploratory search task is the ability to encode and store information collected at every step of the online search as the information need becomes clearer, and this temporary storage of information also enables query reformulation, it can inform the next step of of search process: As the searcher's journey starts with an information need, they need to constantly focus, manipulate ideas, store information for short term manipulation, learn and decide on the next step. In addition, our work in chapter 2 highlighted that one trigger to IO is the brain ability and cognition this would involve the executive functions of the individuals [9]. This is related to the ability to focus, encode and manipulate information. As we have seen in the work reviewed, IO depends on the ability of individuals to store information in working memory [103], in addition individuals who are under IO lack strategies in IR tasks and individuals who are older are more prone to IO [56, 61, 70].

4.1.4 Executive functions and control

Such processes involved in exploratory search and IO are associated with what is denoted as ”executive functions” in cognitive neuroscience. Executive functions constitute a subset of higher order cognitive processes that enable individuals’ to exert cognitive control and goal-directed behaviour, allowing us to regulate our thoughts and actions: The literature agrees on three core cognitive aspects that the executive functions encompass: one is inhibition, which involves self-control and selective attention; the second is cognitive flexibility which involves manipulating ideas and creativity; and finally working memory [242]. In the field of cognitive neuroscience, it has been shown that the prefrontal cortex (PFC), which is situated at the anterior portion of the frontal lobe, is involved in various aspects of executive functions such as working memory [243]. However, additional regions were also found to be associated with performance in tasks involving executive functions. Some of the identified regions through neuroimaging are the frontal and parietal cortex, dorsal anterior cingulate and bilateral anterior insula, and the thalamus [244, 245, 246, 247, 248].

One important executive function paramount to both the search process and IO is working memory: It is a type of short-term memory that is involved in the active manipulation of information and is necessary for cognitive processes such as decision making [249]. It plays an important role in different cognitive functions, such as reasoning and language comprehension [250] and, by extension, a key part of the IR process. In addition to mobilising working memory during the online search process, the searcher has to control their attention and direct it as they progress through their task. In a review of top-down neuromodulation, Gazzaley and Nobre [251] have highlighted the close relationship between working memory and attention where previously it was believed that attention was only involved in working memory up until information was encoded, but they present work which found evidence of the role of attention in working memory maintenance of information.

In the work by Moshfeghi et al. [237] using fMRI in IIR, they identified an activation of the DLPFC when their participants were not in the information need condition, they linked this activation to the role the DLPFC has in working memory: participants used their working memory to retrieve information to provide an answer to their question-answering task. In addition, in the field of cognitive neuroscience, previous work reviewing fMRI based studies has shown that cortical activity in the DLPFC is linked to working memory and manipulation of information in memory [252]. Hence, the main region of interest which we focus on reviewing is the DLPFC which is known to be involved in a wide range of functions, e.g. decision-making [249], learning [253], working memory [252, 254], language comprehension [255], task representation [256], as well as attention [257].

Barbey et al. [254] investigated the role of the left and right DLPFC by using neuropsychological tests (such as the n-back task) on individuals ($N = 19$) who suffered brain damage to the DLPFC. Their study took place between 1967 and 2006 when they acquired structural computed tomography (CT) scans: The left DLPFC was shown to be necessary for the manipulation of verbal and spatial knowledge (e.g letter-number sequencing) in working memory, whereas the right DLPFC was shown to be involved in verbal and spatial reasoning (e.g arithmetic) [254]. In a study by Reuter-Lorenz et al. [258] using positron emission tomography (PET), it was found that DLPFC lateralization was noticeable in younger adults (18-30 years old), as one region had higher activation than the other while when compared to older adults (62-75 years old) both regions had a similar activation: This phenomenon is defined by the model hemispheric asymmetry reduction in older adults (HAROLD), the model is based on neuroimaging studies that show that prefrontal activity is less lateralised in older adults compared to younger adults [259]. Based on the review of this work, we believe cognitive improvement of the left DLPFC could allow better retention in working memory, which would enable better language comprehension and manipulation of information for decision-making during an IR task.

When performing a search task, the searcher is faced with various modalities i.e: text, images, audio, and videos. The searcher has to process the information as part of sense-making. In the case of text, it was shown that the left DLPFC is involved in language comprehension [250, 260].

An important subfield of study in neuroscience is the field looking at modulating activity in the brain. It deals with the stimulation of specific brain regions, some studies have found that it could help tackle post-traumatic stress disorder (PTSD) [261], improve performance on specific tasks, and better understand the brain-behaviour relationship and which brain regions are involved in completing tasks [260]. A non-invasive device which has shown promising results is transcranial direct-current stimulation.

Transcranial direct current stimulation (tDCS) is a non-invasive brain stimulation technique, it modulates cortical excitability and activity, it can facilitate or inhibit behaviour. It is a small portable device that sends a small electrical current using two or more electrodes: target and reference electrodes [262]. In the case of more than two electrodes this approach is called High Definition tDCS (HD-tDCS) [262]. For HD-tDCS the electrode size is smaller (around 4mm in radius) and in the case of two electrodes their area is typically around 35cm^2. A smaller electrode size allows a more focal stimulation effect, while larger electrodes ensure the stimulation of the whole target area. These electrodes constitute the circuit through which the current flows in the brain of the individual. There are two types of tDCS stimulation: Anodal and cathodal. Anodal stimulation depolarises the neurons and is assumed to create facilitory behavioural effects, while cathodal stimulation hyperpolarises the neurons and creates an inhibitory behavioural effect. In order to set up tDCS, one needs to locate the regions of interest and to do so we can use an EEG cap. An EEG cap offers a mapping of the different brain regions based on an agreed standard, and one popular mapping is the 10-20 system, which helps with the placement of the electrodes. The target electrode is put on or around the target region to be stimulated, while the reference electrode can be placed extracephalically (i.e not on the scalp) or on the scalp, and in a region where the stimulation can be maximised, the electrical current movement can be simulated using software like SimNIBS [263]. Prior to placing the electrodes, they are often placed in small, cotton pouches where saline water can be used to facilitate current conduction, or alternatively conductive paste is used rather than saline. Once the electrodes are set up, a specific electrical current intensity and duration are set: Most stimulation lasts between 5 to 30 min, current intensity varies and can go up to 4 mA while maintaining tolerability [264]. While other brain stimulation techniques cause neuronal firing and action potentials, tDCS instead changes the polarity of the membrane and the probability of firing: It is therefore classified as a 'neuromodulation' technique rather than a 'brain stimulation' technique [265]. To ensure that there is no placebo effect, we can incorporate a 'sham' condition: Sham consists of a ramp up period, which usually lasts around 30 seconds, until the current reaches the intensity that was set by the researchers and it is then followed by a ramp down period where the current slowly decreases until tDCS switches off. The sham condition is meant to replicate the sensations of tDCS under actual stimulation and ensure tDCS blinding. Some of the sensations associated with tDCS would be headaches, pain, burning, tingling, and itching.

Neuromodulation, and in this case tDCS has been used to better understand the involvement of brain regions and drive an improvement in performance on various tasks [266], it has also been used to explore its potential benefits for individuals who experienced strokes [267] or suffer from Parkinson's disease, among other clinical conditions [268].

In a study combining the use of tDCS and EEG, Dubreuil-Vall et al. [232] look at the P300 (associated with assessing stimulus value) and N200 (associated with identification of stimulus, selective attention) ERPs: They found that 1 mA anodal tDCS stimulation on the left DLPFC lead to an improvement of the reaction time in a flanker task with an increase in P300 amplitude and decrease in N200 amplitude [232], these results indicate the associated of selective attention with the left DLPFC. Using a brain stimulation technique, namely repetitive transcranial magnetic stimulation (rTMS), researchers showed a causal relationship between the left DLPFC and perceptual decision-making: Philiastides et al. [269] used 1 Hz low-frequency rTMS stimulation for 12 min on the left DLPFC and asked participants to categorise stimuli, stimuli presented at a fast pace, were able to show that it lead to lower accuracy, drift rate (rate of information accumulation) and faster response time in active stimulation relative to the sham condition [269]. In a double-blind tDCS study with sham as a control condition, Klaus and Schutter [260] applied for 20 to 30 min a 2mA cathodal tDCS stimulation where the cathode was placed on the left DLPFC and the anode positioned between the Fz and

Cz region. Participants had to match a sentence to two illustrations as stimuli, they found that the participants had a longer reaction time in the cathodal tDCS stimulation compared to the sham tDCS stimulation. These results support the involvement of the left DLPFC in language comprehension and complement prior studies which found that anodal tDCS stimulation on the left DLPFC improved reading performance [250], it also indicates that language comprehension is linked to working memory. As working memory is defined as an active loop where information is stored temporarily, language comprehension and production would not exist without it. Working memory is assumed to have both processing and storage capabilities, Daneman and Merikle [270] in their meta-analysis of working memory and language comprehension, they show the processing capability of individuals combined with working memory capacity is a better predictor of language comprehension than working memory capacity alone: The predictor tasks used for verbal (e.g listening span for processing and storage, word span for storage alone) and maths (e.g. counting span for processing and storage, digit span for storage alone).

Cerruti and Schlaug [271] showed that a 1 mA anodal tDCS stimulation of the left DLPFC improved performance on a compound remote associates task when compared to anodal stimulation on the right DLPFC [271]. In another study, anodal tDCS stimulation was administered for 11 minutes at a 1 mA current on the left DLPFC, and it was found to improve solution recognition for difficult problems with greater performance improvements in participants with greater trait motivation [272]: Their task was a variation of the compound remote associates task, but with an emphasis on solution recognition as they presented the problem for a shorter duration and followed it directly with the target word, then participants had to specify if it was the solution to the problem or not — they used the BIS/BAS questionnaire to measure trait motivation [273]. This shows the potential to improve the users' interaction with IR systems when trying to encode new information and help the searchers identifying the relevance of search results during their exploratory search.

Miler et al. [274] investigated the effect of 2 mA anodal tDCS stimulation on the PFC and its impact on attention networks. They designed a double-blind study and the stimulation lasted for 20 minutes: They used the Attention Network Test (ANT) to test the effect of tDCS on executive attention, they found that 2 mA anodal tDCS stimulation resulted in greater executive attention based on Attention Network Test compared to sham tDCS stimulation. The ANT is a popular tool which consists of a modified Flanker test and measures alertness, orienting (ability to pay attention to a stimuli), and executive control, more specifically conflict-resolution (ability to ignore a distractor stimuli) [275]. The ANT has been used in various studies looking at attention disorders [276], factors which can lead to improved attention [277], as well as identifying the relationship between executive functions [278]. In the work by Redick and Engle [278], they found using the ANT that working memory capacity impacted executive control as the participants of their study with lower working memory capacity had a larger reaction time when facing incongruent stimuli compared to participants with larger working memory capacity. However, they did not find any difference on the other two indexes of the ANT, i.e alerting and orienting.

Although neuroscience has done extensive work to study the impact of brain stimulation on the left DLPFC and right DLPFC, all the tasks were taken from the field of neuroscience and were oriented to tasks that were not representative of real-life scenarios such as IR tasks. This motivates us to build on the previous work, and design an experiment where we investigate the impact of tDCS on an exploratory search task. In Information Systems, a field which studies the development of information and communication systems in organisations and societies, we find Neuro-Information Sytems (NeuroIS), it studies Information Systems using neurophysiological research tools [227]: They used non-invasive brain stimulation in an attempt to better understand the role of the left and right DLPFC in the perceived ease of use of an e-commerce system: They used anodal and cathodal tDCS stimulation on both left and right DLPFC, however, their study did not find any significant change in perceived ease of use, but did find that participants under anodal on the left DLPFC and cathodal on the right DLPFC saw a reduction on the fixation time spent on the buy buttons, it could be that their participants were more prone to a less conservative decision-making [266].

Neuromodulation has yet to be explored by the IR community. The advantage of using this technique is that it allows the design of realistic IR tasks and not only test the role of specific brain region in its accomplishment, but also explore the application of non-invasive brain stimulation in improving users' experience with IR systems. Until now, the tasks which explored the cognitive manifestations and associations in IR were limited due to the type of neurophysiological data collection methods used: In the case of MRI/fMRI, participants of a study had to enter a scanner while lying down with limited control over a screen. This approach is useful when controlling for a specific independent variable and wanting to compare the brain regions activated and activation levels, however, as we have highlighted before it does not have good temporal resolution. The other approach is EEG, which relies on capturing brain signals, which are elicited responses of the brain to an event. EEG offers more flexibility than fMRI as the participants are sitting on a chair and can interact directly with a computer, it also offers temporal accuracy in the order of milliseconds with regard to brain activity. Although mobile EEG is a growing field, most of the past studies we reviewed in IR/IIR used traditional, static EEG, which is extremely sensitive and requires the participants to stay still and focused during the task in order to not degrade signal quality.

In the present work, we leveraged tDCS to neuromodulate the left DLPFC. As we have highlighted throughout, the importance of one key component of whether IO or exploratory search, which is working memory, we have also shown how working memory is strongly associated with additional cognitive processes such as attention and decision-making. We have also seen that all these cognitive components fall under what is denoted as executive functions. One established region of the brain for executive functions that we identified is the PFC. Finally, we have seen that the left DLPFC is strongly linked to all the various executive functions we discussed that were related to the exploratory search process as well as IO. All these elements motivated our work and lead us to ask the following research questions:

- **RQ 7**: Can we reduce the dwell time of individuals on an online search task by applying transcranial direct current stimulation on the left dorsolateral prefrontal cortex ?

- **RQ 8**: What is the impact of the left dorsolateral prefrontal cortex stimulation on the interaction of individuals with an information retrieval system?

To the best of our knowledge, this is the first study of this kind in IR and IIR to look at the impact of anodal tDCS stimulation in an exploratory search task and investigate its impact on the dwell time (time spent on web pages excluding the time spent on the search engine home and results' web pages).

We chose 2 mA of anodal tDCS stimulation. A similar current intensity was found to lead to improved attention in the work by Miler et al. [274] where they used 2 mA anodal tDCS stimulation on the left DLPFC. Our anodal stimulation was on for a total of 15 minutes with a filler task (reaction time task) which lasted 8 minutes followed by the exploratory search task which also lasted 8 minutes. Having a filler task ensured that the anodal tDCS stimulation effect had started.

We hypothesised that the dwell time would be reduced because individuals with improved executive function will be able to store more information in working memory, have facilitated language comprehension, better executive attention and therefore should be able to process and filter information faster (RQ 7). We pre-registered our design and hypotheses on the 'Open Science Framework' platform (subsection 4.2.9). In addition, we performed an exploratory analysis looking at additional outcome measures which we believed could be impacted by stimulation of the left DLPFC and the executive functions associated with it (RQ 8).

4.2 Methodology

4.2.1 Ethics

Ethics approval was given by the University of Glasgow, College of Science and Engineering Ethics Committee. Upon their arrival to the lab for the study, the participants were given a safety questionnaire [279]. We screened participants using our tDCS safety questionnaire based on the following: (1) An adverse reaction to tDCS in the past; (2) a seizure; (3) an unexplained loss of consciousness; (4) any brain-related or neurological injury; (5) metal in the head; (6) implanted medical devices; (7) frequent headaches; (8) specific medications (more details in the appendix); (9) recent consumption of recreational drugs or alcohol; (10) being sleep deprived; (11) pregnancy; (12) epilepsy in the family (Appendix B.2); if participants answered yes to any of the questions, except whether they used tDCS in the past or wanted to find out more, then participants weren't allowed to take part to the study. Once they were screened and seen fit to do the study, we provided them with the information sheet (Appendix B.1), and the consent form (Appendix B.3). In the information sheet, the participants were provided with detailed information on the whole procedure: They were made aware of the potential side effects of tDCS and that the current was increased gradually to minimise the potential of these side effects happening. In the information sheet, we specified in bold that if the participants felt any of the following symptoms, they should remove the electrodes from their scalp and inform the researcher: (1) altered vision; (2) lightheadedness; (3) dizziness; (4) involuntary movements such as eye or muscle twitching; (5) confusion; (6) nausea; (7) loss of awareness; (8) convulsions; (9) cramps; (10) disorientation; (11) motion sickness; (12) Discomfort or pain in the head or eyes; (13) Any other symptoms that you would class as atypical or unusual. All participants provided informed consent at the start of their first session. We informed the participants that their participation is voluntary, and that they were free to withdraw at any time without providing any advanced notice or an explanation. All the data collected was anonymised and each participant was assigned a randomised user ID for the study. The data was stored in a secured location (password protected). The study was run in accordance with the Declaration of Helsinki.

4.2.2 Participants

A simulated power analysis (with Cohen's d=0.45 [280], alpha=0.05 and power=0.8) identified that 32 participants were required to perform a one-tailed, paired sample t-test to compare the change in dwell time between the active and sham tDCS conditions. We recruited 32 participants with a mean age of 26 (range=19-41, SD=4.7, 17 males, 14 females, and 1 preferred not to disclose) who met the inclusion criteria. Written, informed consent was obtained from each participant at the start of session 1,

(Appendix B.3).

Participants demographics

Through our entry questionnaire, we asked participants to self-report their English level, share their countries of citizenship, and share their education level: We had 29 participants fluent in English, 2 intermediate and 1 had basic competence in English. We had 19 different countries of origin, our most represented countries were the UK (23.7%), Albania (12.5%), followed by Ghana (8.7%), and China (5%): The rest of the countries were across Europe, Africa, and Asia. Most of our participants had at least a Bachelor's degree (46.3%), or master's degree (31.2%), and the rest either had a high school degree or some college but no degree. We report on the countries of origin and the education level as these are important to highlight that we are not creating any bias with our recruitment, we aimed to have a diverse pool of participants as information can be searched in a different way by people with different backgrounds or due to their education level which can lead to unequal information literacy search skills [281].

Across the 32 participants, 2 participants were left handed and 30 right handed as per the Edinburgh Handedness Inventory (EHI) [282].

Search engine familiarity

In our entry questionnaire (Appendix B.4), we asked participants to self-report on their proficiency with search engines, which search engines they used, and how often they used search engines: 9 believed they had excellent search engine proficiency, 19 believed were good and 4 thought they were average. We found that the most commonly used search engine by our participants is Google followed by Bing (figure 4.2).

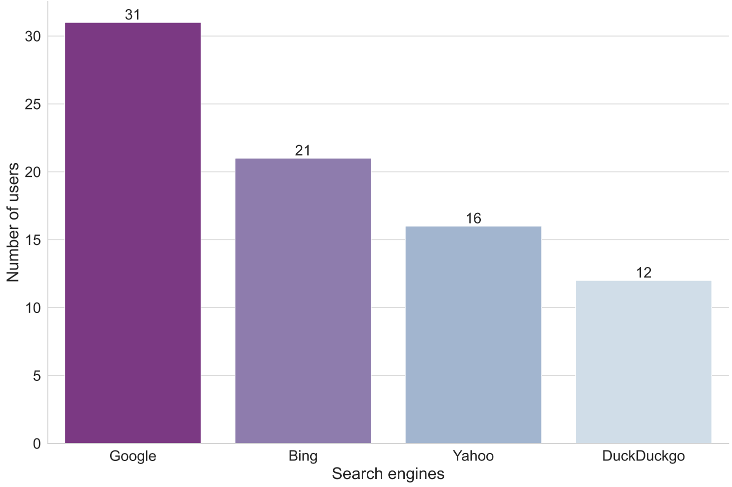


Figure 4.2: Search engines previously used by our participants

Except for 1 participant who reported that they use search engines every few days per week, the rest of the participants use search engines every day.

4.2.3 Study design

The study used a within-subject design, and it was a randomised, double blinded experiment. The independent variable was the tDCS condition. The participants came for two sessions: in one they had to perform the search task during sham tDCS and the other during the active anodal stimulation. We collected behavioural data using an in-house search engine developed for the purpose of the experiment. The search engine was designed to resemble a widely used and popular one. The data collected comprised the list of websites accessed, queries submitted (only the queries typed in the search bar and submitted), mouse data related to which search result they hovered on, as well as the associated timestamps for each event. Each of the 2 experimental sessions lasted 60-75 minutes, the time varied since participants had to fill out questionnaires and there was also the time we had to take to set up the tDCS' electrodes.

4.2.4 Experimental procedure

Participants came for two sessions in a repeated measures design. Ahead of each session, participants had to fill in a tDCS safety questionnaire (Appendix B.2). The questionnaire served to ensure the participants were eligible to have tDCS stimulation. At the beginning of the sessions, the participants were provided with the experiment information sheet. After they signed a written consent form, the tDCS electrodes were attached to the scalp. Once the electrodes setup was done, the participants had to spend 5 minutes familiarising themselves with the search system's interface. Participants then selected two topics for the two exploratory search tasks and completed a pre-task questionnaire where they explained the rationale for their choice. They were then presented with the topics they selected in a randomised order through the survey platform we used.

The participants had to then perform the first exploratory search task which represents either base sham or base active condition depending on the condition order (Block 1 in Figure 4.3). At the end of the first search task, the participants completed a post-task questionnaire, then they were given a practice reaction time (RT) task. After the practice RT task, the experimenter initiated the tDCS protocol, at which point the participants began the RT task (Block 2 in Figure 4.3). At the initiation of the tDCS ramp-up, participants performed the reaction time task for 8 minutes (the data was not analysed as this was done to ensure participants remained focused throughout the experiment). At the end of the reaction time task, the participants immediately performed the second exploratory search task for 8 minutes; this was either sham or active condition (Block 3 in Figure 4.3). At the end of each session, participants completed a post-task questionnaire to assess their experience with Moogle and provided an unbiased overview of the topic chosen.

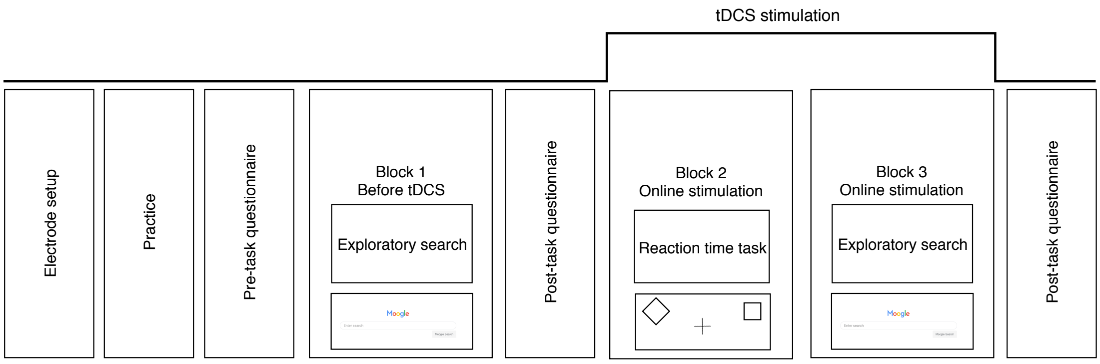


Figure 4.3: Experiment design

4.2.5 Transcranial direct current stimulation

A direct current was applied using a battery-driven constant current simulator (NeuroConn GmbH, Germany). Two tDCS protocols were applied in a double-blinded, counterbalanced, repeated measures design with a minimum of 48 hours between sessions. In the anodal condition, a direct current of 2 mA was delivered for 15 minutes and in the sham condition a 2 mA current was applied for 20 seconds. In both protocols there was an additional 30s ramp-up period at the start of stimulation, where the current gradually increased to 2 mA, and a 30s ramp-down period at the end of stimulation. In both conditions a 5x5cm anode was centred on the left DLPFC (EEG electrode location AF3) and a 5x7cm reference electrode (cathode) was placed just anterior to the vertex (Cz). The electrode positions were identified using SimNIBS [263] to optimally target the left DLPFC as illustrated in figure 4.4. The electrodes were composed of carbon rubber encased in 0.9% saline soaked sponges and held in place using an elastic headband. The "study mode" of the NeuroConn device was used to double-blind the stimulation condition to both the experimenter and the participant. The experimenter initiated either the active or sham protocols by entering a 5-digit code from a list which had been prepared in advance by a collaborator (who took no part in the participant testing). The screen of the NeuroConn device provided real-time resistance readings for both protocols, meaning that the experimenter could not identify which of the two protocols had been initiated.

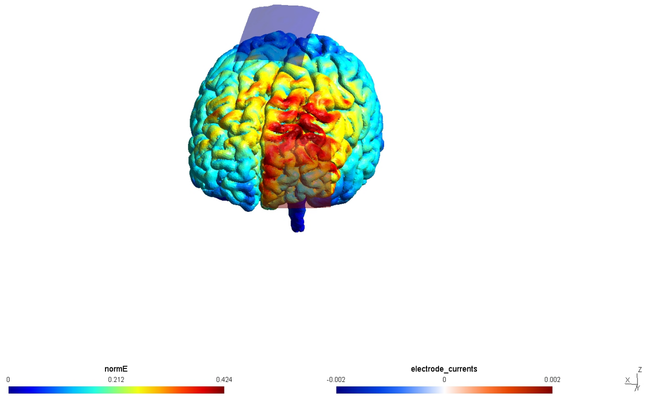


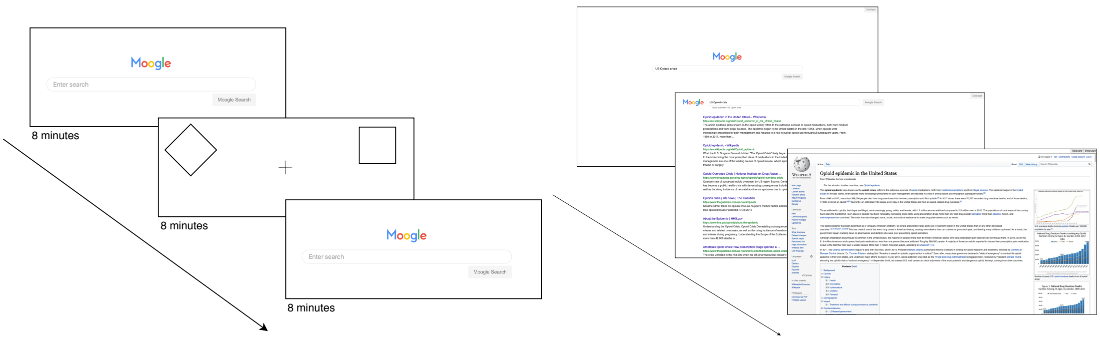
Figure 4.4: Current intensity is the strongest on the left DLPFC when stimulating AF3 and the reference electrode on Cz

4.2.6 Task description

For the purpose of the experiment, we built an exploratory search dataset which consisted of 7 topics (Appendix B.8): (1) China's social credit system, (2) cognitive biases, (3) WhatsApp security, (4) US Opioid crisis, (5) Huawei and the USA, (6) Sudan after December 2018 protests, (7) Alternatives to meat.

The topics were carefully selected and covered a wide range of areas. We picked them from the news of the months leading up to the experiment. We made sure they were popular enough to generate interest but also to make sure not many people would have an in-depth knowledge. This was done to generate interest and increase engagement from the participants, and ensure the participants would find a few topics which they are unfamiliar with. The topics were chosen to be open-ended, and so they would create a fuzzy information need and to unsure an iterative search process: these elements are all key to an exploratory search task [210]. In addition to the basic criteria for a good exploratory search task, we also wanted to have topics which would minimise the risk of emotional reactions. The order in which the topics were displayed was randomised and each topic had a topic ID assigned for reference, the same 7 topics were displayed across the first and second day of the study, and they had to choose 2 topics for each session. In each topic participants were presented with a hypothetical scenario, e.g "Your colleague tells you about the recent WhatsApp vulnerability. Using the search engine provided, find out about the origin of the hack and data collected so you can convince your friends of the risks with using instant messaging apps". The aim of these scenarios was to gather as much information as possible in order to provide an overview of the topic.

Participants were first given time to familiarise themselves with the search system. Once they were accustomed with the search system, we asked them to perform the first search task on the first topic chosen. The experiment consisted of three blocks: Block 1 was the first search task, block 2 was the reaction time task, and block 3 was the second search task. Each block was performed for a fixed duration of 8 minutes. Prior to the beginning of the first block the participants chose the two topics part of exploratory search. During the first block (search task 1), they were asked to gather as much information as possible through Moogle. While they performed the task, we collected data on the various websites visited and the associated timestamps for subsequent analysis. At the end of the task, they were asked to provide an unbiased overview of the topic on an online form. Following the first block, the stimulation was initiated and a reaction time (RT) task was performed (Figure 4.5). Once the reaction task ended, the participants transitioned directly to the third and final block (search task 2). Once again, at the end of the search task, they were asked to provide an unbiased overview of the topic. The total dwell time was calculated for search task 1 (before tDCS, *baseline*) and search task 2 (after tDCS) and used in the statistical analysis.



(a) Tasks (b) Example of a search process

Figure 4.5: The different tasks of the experiment with an example of a search task

4.2.7 Search system: Moogle

Moogle is our search engine which is designed to replicate the same experience users would have with Google. To encourage efforts to perform additional studies in similar conditions, Moogle is available at: https://github.com/aminecs/moogle. It was designed to represent as closely as possible an experience similar to the one of a mainstream search engine. It offers us the possibility to collect interaction data of the users with a search engine and it allows us to extend it with features such as relevance judgment. We set up Moogle to run locally and we used a Macbook Pro laptop connected to a DELL LCD monitor, they performed the tasks on the DELL LCD monitor and didn't have access to the laptop's screen. We also used an external keyboard and mouse for the participants to perform the task.

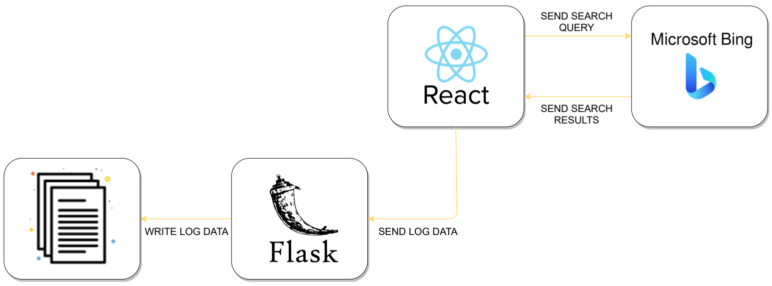


Figure 4.6: Moogle architecture

For the development of Moogle, we had three main components which communicate with each other (figure 4.6). The user facing component was developed in React, when a user submitted a query, we made a GET request to the Microsoft Bing API and retrieved the results. The results displayed were formatted in a traditional fashion on the search engine result page (SERP) using the JSON data returned by the API call. In addition, the search engine results page was always in a single column format, we did not have a knowledge panel implemented (Figure 4.7). Once participants clicked on a search result, they did not leave our system instead they were shown the new web page through an iFrame, and this allowed us to have an explicit relevance feedback mechanism: The only difference with the traditional experience participants would have with Google is that we had explicit relevance feedback buttons ('Relevant'/'Irrelevant') at the top left of the web page they were reading. This explicit relevance feedback mechanism allowed us to collect the exact time (timestamps) participants were done with a specific search result and were returning to the search results page.

The cache and cookies were cleared after each participant's session, this was done to ensure that there were no recommended queries or hyperlinks highlighted in the search engine results page (i.e purple colour which indicates if someone clicked on a specific result).

While the user was interacting with Moogle, the React app made various POST requests containing interaction data to the Flask app which wrote it out locally in comma-separated files.

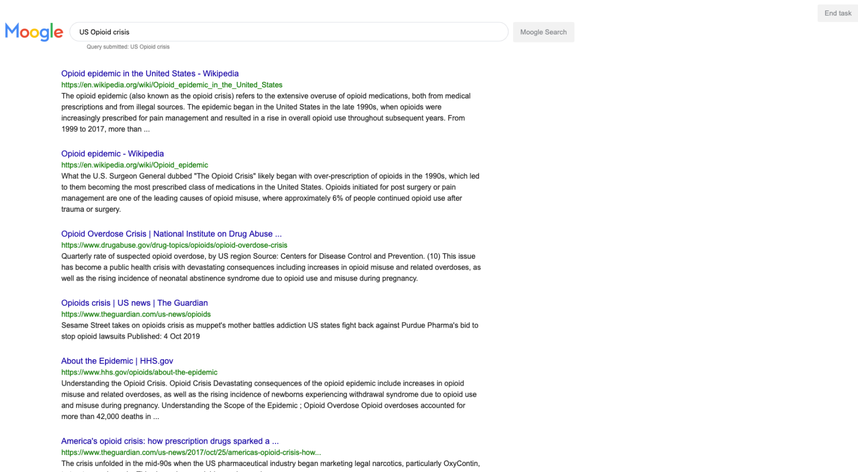


Figure 4.7: SERP of Moogle

The code is available at `https://github.com/aminecs/moogle`

4.2.8 Reaction time task

We believed that the effect of tDCS stimulation would not start as soon as stimulation started, therefore we used a 'filler' task during the first half (8 minutes) of stimulation. A reaction time (RT) task was designed as to be our filler task which would allow our participants to remain engaged with our study as they start getting stimulated. Participants had to perform a right or left click as fast as possible depending on whether they observed in the middle of the screen a square (right click) or a diamond (left click).

Between each stimuli, a fixation cross was displayed in center of the screen for a period varying between 1700ms and 2100ms. This task lasted for 8 minutes.

The RT task was built as a web application using React and we used the same Flask based backend that powered Moogle but used a specific endpoint for our POST requests (/postReactionTimeStats) to record the response of the participants for every stimuli as well as the associated timestamps. This allowed us to not have an additional microservice to manage and we only had to have one backend service running for the whole duration of the study.

4.2.9 OSF pre-registration

Once the design of our experiment was finalised, we pre-registered it on the 'Open Science Framework' (OSF) platform along with the questionnaires used (Appendix B). In our design, we hypothesise that 15min of anodal stimulation will result in a reduction of the dwell time, i.e the amount of time spent on the websites visited during a search task (excluding the time spent on the search engine) will be smaller under active stimulation.

In our statistical analysis section, we define how we analyse the dwell time: To evaluate whether the reduction has happened, we define ΔDT as the total dwell time in the baseline search task subtracted from the active/sham total dwell time. We highlighted that we would also do a further exploratory analysis by looking at different outcome measures which might be subject to the brains executive functions.
The OSF pre-registration is available at: `https://osf.io/mj4dp/`

4.2.10 Outcome measures

In our pre-registered protocol, we defined dwell time as our main dependent variable: The dwell time is the time spent on an article; it starts from the moment they click on a result and leave the search engine result page and ends the moment they click on one of the relevance feedback buttons (*Relevant/Irrelevant*).

In addition, for our exploratory analysis, we looked at various additional outcome measures to see whether tDCS had an impact on these. These additional outcome measures are: SERP dwell time, the number of queries, formulation time of queries, similarity of queries, number of websites visited, number of unique websites visited, number of urls hovered on, and number of unique urls hovered on. Below we define the outcome measures which might not be clear from their name:

Search engine result page (SERP) dwell time

The SERP dwell time represents the total time spent by a participant on the search engine results pages during a task.

Formulation Time of Queries:

We calculated the total time by summing up the time taken to write each query divided by their number of characters.

Similarity of Queries :

Given an exploratory search task, we calculated a similarity score between the queries formulated and calculate the mean. To calculate a similarity score, we generated a sentence embedding for each query using the library 'HuggingFace' and the model 'all-MiniLM-L6-v2'[2] [283], each sentence is mapped to a vector of 384 dimensions. Sentence embeddings are a vector representation of a sentence that is generated using a pre-trained machine learning model. The model is trained on a large a corpus of text and at the end, depending on how it was trained, it is able to create a vector representation for words, sentences and even documents: This is called the inference step, based on an input it will effectively create an encoding in vector format. Since these large language models are trained on large corpus of text, all the words of the English vocabulary will have a vector representation.

We use the 'HuggingFace' inference API to generate the embeddings for each query and compute a similarity score. The implementation calculates a cosine similarity score between the embeddings and this score is between 0 (not similar at all) and 1 (it's the same sentence). This approach allows us to determine whether 2 sentences are semantically similar.

When a participant formulated only one query during a search task, we assigned a score of 1.

[2]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2, last accessed: 31/07/2022

URLs hovered on

These are the search results which the participants hovered on when they were on the search results page (Figure 4.7).

4.2.11 Statistical analysis

Before performing any statistical significance test, we checked whether our data was normally distributed using Shapiro-Wilk test [177]. For parametric data, we used repeated measures t-tests. For non-parametric data, we use Wilcoxon signed-rank tests. We set α to 0.05. In the explanatory analysis of our work, we used linear mixed-models (LMM) in order to integrate various fixed effects and random effects. To fit our models, we used Restricted maximum likelihood (REML) rather than the Maximum likelihood method, the latter was shown to introduce bias on small sample sizes [284]. We used LMMs to evaluate the behaviour of our fixed effects, i.e whether the dependant variable increases or decreases given a variable, in this condition or modulating factors. In addition, as described by Brown [285] to evaluate whether a specific variable had statistically significant effect, we used the likelihood-ratio test and had to refit our models using the Maximum likelihood method since we had a different number of fixed effects across the models. When building these LMMs, we converted any time data from milliseconds to seconds. Our data analysis was performed using Python and R, to build LMMs we used the `lmer` function from the `lme4` library and visualisations such as Rainclouds [286] were made in RStudio.

4.2.12 Questionnaires

At the beginning of the experiment, participants were given an *entry questionnaire* (Appendix B.4) to collect demographic information. We also ask how often they use search engines and which search engines they use. Before the first search task, we asked the participants to fill in a *pre-task questionnaire* (Appendix B.5) which asked about the topics they had chosen and why they made those choices. We also gathered information about their perception of the tasks they were about to perform. We asked them about the tasks in terms of difficulty, stress perception, interest, how clear they thought the task was and how familiar were they with it, for each of these we used a 5-point Likert scale for the answer from 'Strongly disagree' to 'Strongly agree'. After each search task, the participants were asked to fill out a *post-task questionnaire* (Appendix B.6). They were asked to provide on the survey platform an in-writing an answer to the topic they have looked up and how much more they thought they learned about the topic *(A great deal/A lot/A moderate amount/A little/None at all).* We also asked them about their experience with Moogle: *How was your experience with Moogle during this task? (answer: Extremely satisfied/Somewhat satisfied/Neither satisfied nor dissatisfied/Somewhat dissatisfied/Extremely dissatisfied).* Similarly to the pre-task questionnaire, we asked them about their perception of the task. Finally, at the end of the session, participants filled an *exit questionnaire* (Appendix B.7). We asked them to rate on a scale (*answer: Not at all/Slightly/Moderately/Very/Extremely*) whether they experienced side-effects during the session *(headaches, tingling, itching, burning, pain)* and how long they felt they lasted for (*answer: Not applicable/A few seconds/1-2 mins/3-5 mins/6-10 mins/11-15 mins*). At the end of the second session, we asked the participants to guess which session was inactive (sham), how certain they were on a scale from 0 to 10, and their reasoning.

BIS/BAS questionnaire

The participants had also to fill out the BIS/BAS questionnaire [273] aiming at measuring their trait motivation score. In a previous study with tDCS [272], it was shown that trait motivation had a role in the performance changes: Metuki et al. [272] in a study where they stimulated the left DLPFC and looked at how tDCS impacted problem solving, they found that participants with a lower BAS score benefited more from brain stimulation as they improved in problem solving for difficult items.

One measure for trait motivation which is used is the BIS/BAS questionnaire [273]. It is a 24-item questionnaire, each item's response consists of a 4-point Likert scale from 1 (strong agreement) to 4 (strong disagreement) with no neutral response. The questionnaire consists of 2 scales: Behavioral inhibition system (BIS) and behavioural activation system (BAS). The BIS scale consists of 7 items and provides an insight into the reaction of individuals to punishment. The BAS scale is split into 3 subscales: BAS drive, BAS fun seeking, and BAS reward. The BAS drive measures the persistence to pursue a desired goal, the BAS fun seeking measures the desire for new rewards and interest in pursuing new rewards in the spur of the moment. The BAS reward scale indicates the response or anticipation to a reward (a high BAS reward reponsiveness score means lower sensitivity to reward, and inversely a low score means higher sensitivity to reward).

4.3 Results

4.3.1 Descriptive results

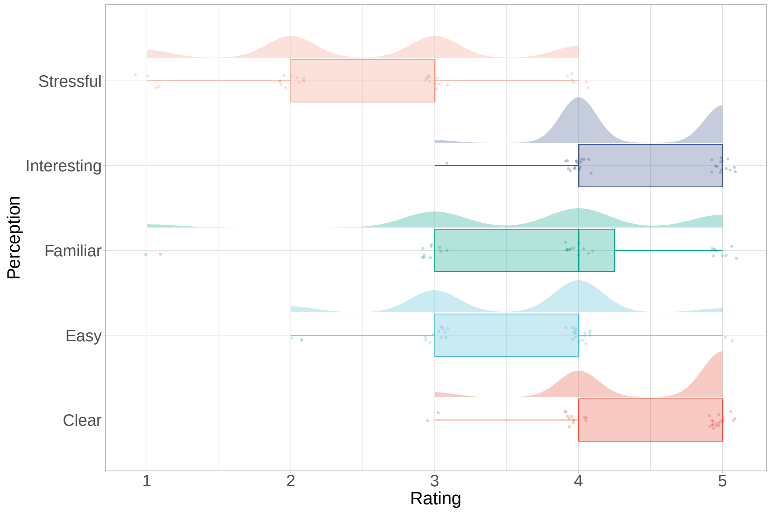
In this section, we report results from our pre-task and post-task questionnaires (subsection 4.2.12): Participants had to self-report their experience with our search engine, and the pre-task perception questionnaire and the post-task questionnaire were meant to give us an indication of how the participants perceived the task before and after.

Following the descriptive results from our pre-task and post-task questionnaires, we present the results from the BIS/BAS questionnaire, then we look at how side effects and correct condition guess (i.e whether participants correctly guessed which session was sham) varied across our participants.

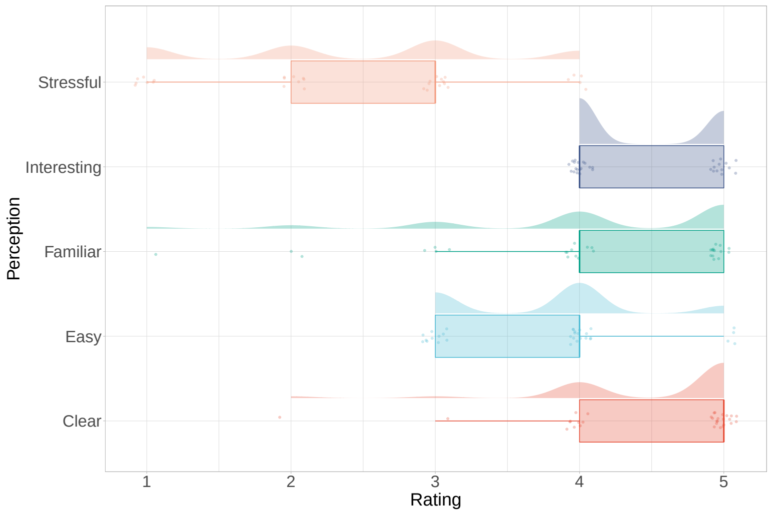
4.3.1.1 Pre-task perception

Due to an issue with the survey collection website that was used, we were unable to retrieve the pre-task questionnaire result for 1 participant (p=13) for their second session, and therefore we decided to drop this participant's pre-task perception from this analysis.

The pre-task questionnaire was given to the participants after they completed the entry questionnaire, we evaluated their perception of the task (*"The task we asked you to perform is going to be easy/stressful/interesting/clear/familiar"*) based on 5-point Likert scale.



(a) Pre-task perception rating from the start session 1, completed prior to the first task.



(b) Pre-task perception rating from the start of session 2, completed prior to the first task.

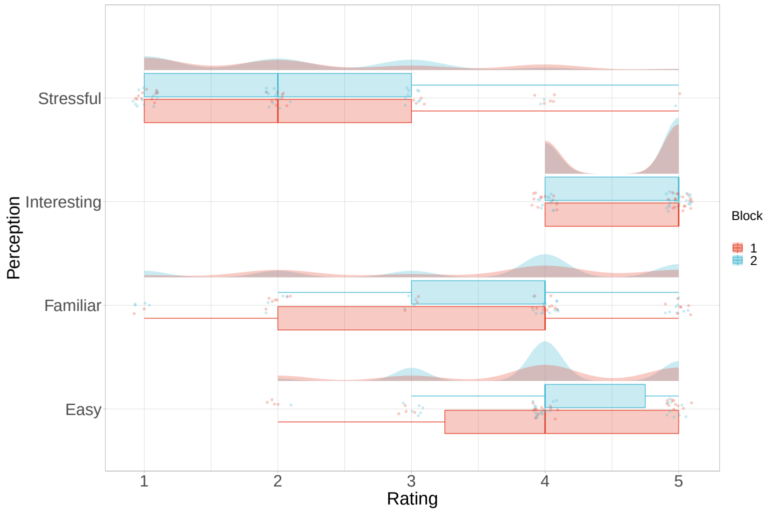
Figure 4.8: Pre-task perception ratings across session 1 and 2.

In Figure 4.8, we can see that on our 5-point Likert scale (*Strongly agree/Somewhat agree/Neither agree nor disagree/Somewhat disagree/Strongly disagree*), participants across both sessions before performing the first search task agreed that it was clear ($Mdn = 5$, $IQR = 1$) and interesting ($Mdn = 4$, $IQR = 1$). The change across the session happens with the familiarity of the task where participants report they are more familiar to it before session 2 ($Mdn = 4$, $IQR = 1$) compared to session 1 ($Mdn = 4$, $IQR = 1.5$). This is coherent as the participants had already done the task once since we had a within-subject design.

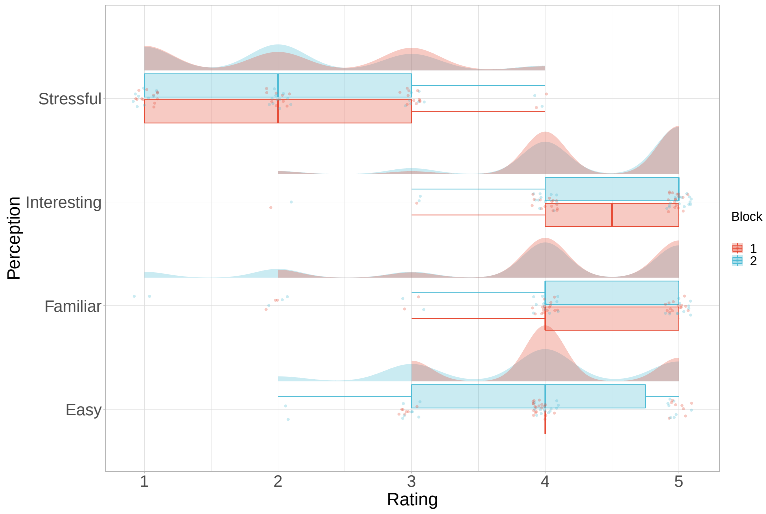
Another change across sessions happened where the minimum value for how easy the task is perceived to be increased to 3 (Figure 4.8b) i.e no one disagreed with the task being easy, and the median remained the same ($Mdn = 4$, $IQR = 1$).

4.3.1.2 Post-task perception

In Figure 4.9, we denote the order in which each participant did the search tasks by ‘block’. The first block represents the search task when they were not under anodal or sham tDCS stimulation (i.e control task), while the second block consists of when the participants were either under active or sham tDCS stimulation.



(a) Post-task perception rating during session 1 across block 1 and 2.



(b) Post-task perception rating during session 2 across block 1 and 2.

Figure 4.9: Post-task perception ratings across session 1 and 2. Block 1 represents our first task when the participants were not under stimulation (i.e base active or base sham), block 2 represents when participants were under tDCS stimulation (i.e active or sham)

A similar observation to the pre-task questionnaire, we find that the data for the participants' familiarity was less dispersed after performing the search tasks during the second session ($Mdn = 4$, $IQR = 1$) compared to the first session ($Mdn = 4$, $IQR = 2$): There is more consensus on the perception that the task was familiar. Finally, we can also note that the search tasks were not necessarily perceived to be stressful and overall, regardless of the session and the order in which the task was performed, the participants perceived their experience across the tasks to be mostly similar.

Experience with Moogle

After each task, we collected information about the participants' satisfaction with Moogle ("How was your experience with Moogle?") using a 5-point Likert scale.

In total, across the 128 tasks performed, 51% of the time participants were *"Somewhat satisfied"* with their experience with Moogle, 45% were *"Extremely satisfied"*, 2% were *"Neither satisfied nor dissatisfied"* and 2% were *"Somewhat dissatisfied"* (figure 4.10).

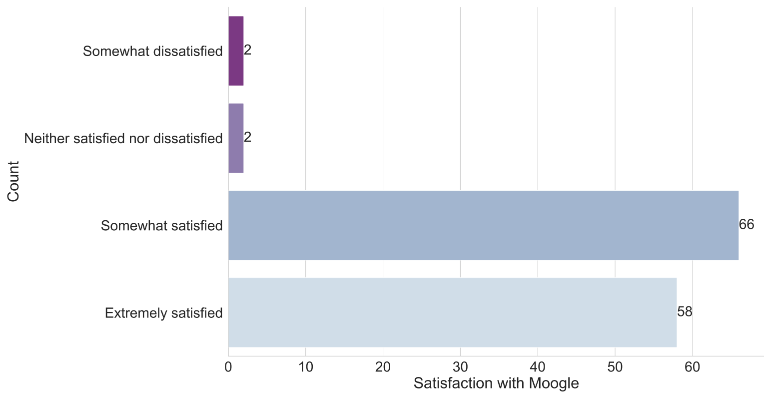
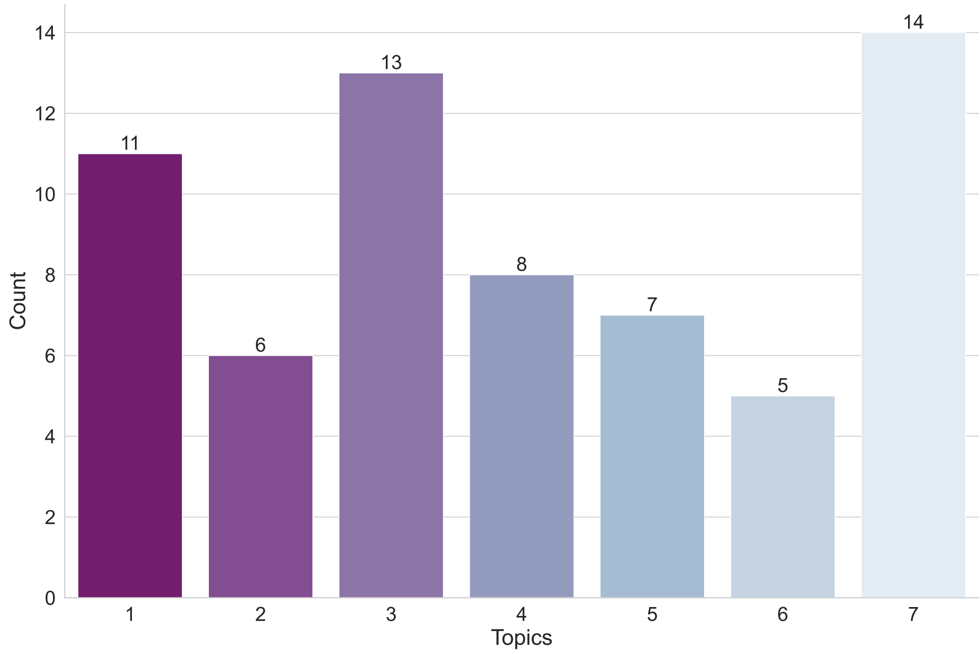


Figure 4.10: Moogle experience satisfaction

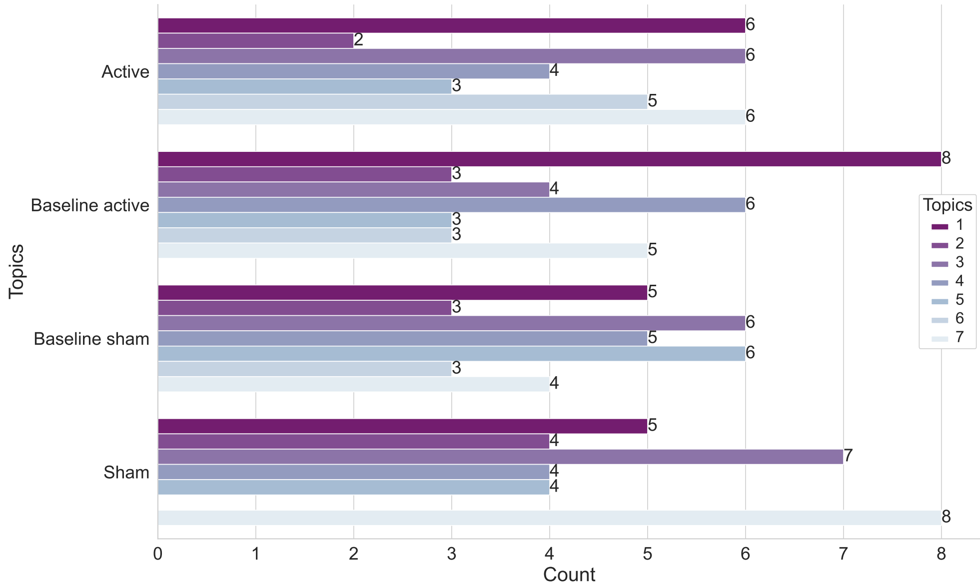
4.3.1.3 Topics

Our topics were diverse and participants were asked to choose the topics they knew the least about and found the most interesting.

We found that the topics were not chosen evenly and some topics were selected more than others (Figure 4.11a). Topic 1, 3, and 7 were chosen most often while topic 6 (Sudan after the December 2018 protests) was the least popular, i.e topic 6 was not chosen by a large number of participants. When we further analysed the topic choices across tDCS conditions, we found that no one performed topic 6 during sham stimulation (Figure 4.11b).



(a) Topics chosen



(b) Topics chosen by participants with a breakdown across conditions

Figure 4.11: Topics chosen by participants and the breakdown across conditions

4.3.1.4 BIS/BAS score

In this section, we report the results of the BIS/BAS questionnaire. Previous work used the BAS score to see whether it moderated behavioural effects of tDCS [272], however, we also report the BIS score as we tested for the BIS and BAS scores as moderating factors.

Our participants had a mean of 24.31 (out of 52) on the BAS scale and 14.31 (out of 28) on the BIS scale (table 4.1).

The division between male and female participants was fairly well distributed (N=32, n_f=14, n_m=17, 1 undisclosed).

		BIS	**BAS**			
			Fun seeking	Drive	Reward responsiveness	Total
Female	Mean	**12.50**	8.71	8.57	**7.64**	24.93
	STD	4.10	2.92	1.83	1.78	5.25
Male	Mean	15.76	**8.58**	**7.82**	7.70	**24.11**
	STD	3.32	1.73	2.62	1.92	4.91
All	Mean	14.31	8.59	8.13	7.59	24.31
	STD	3.93	2.28	2.27	1.86	5.01

Table 4.1: Summary statistics of BIS/BAS results

4.3.1.5 Side effects

At the end of the tDCS stimulation after each session, we asked the participants to report any side effect they might have felt as well as how long they felt they lasted. They could rate the intensity and duration of five side effects: Headache, tingling, itching, burning, and pain.

During the sham condition, 24 participants felt a side effect, while during the active condition 26 participants felt a side effect.

Side effect type	**Sham**	**Active**
Burning	14	12
Itching	18	18
Tingling	19	22
Pain	7	4
Headache	5	4

Table 4.2: Breakdown of the number of times side effects were felt across each condition

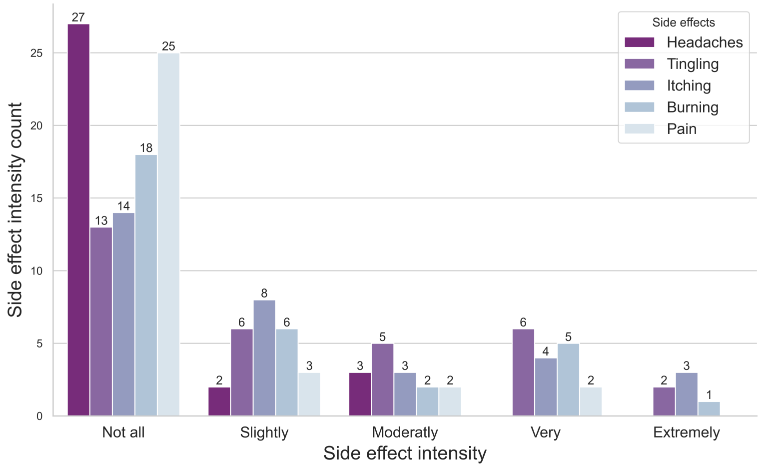
In Table 4.2, we find the most common side effects are itching and tingling. We have asked the participants to grade the intensity of the side effects from 0 (not at all) to 4 (extremely). When looking at the median intensity for each type of side effect (Table 4.3), we find the median tingling intensity was higher in active (Mdn = 1.5, IQR = 2.25) than in sham (Mdn = 1, IQR = 2), while the itching median was the same (Mdn = 1, IQR = 2) across conditions, p = .72.

Side effect type	**Sham**	**Active**
Burning	0 (1.25)	0 (2)
Itching	1 (2)	1 (2)
Tingling	1 (2.25)	1.5 (2.0)
Headache	0 (0)	0 (0)
Pain	0 (0)	0 (0)

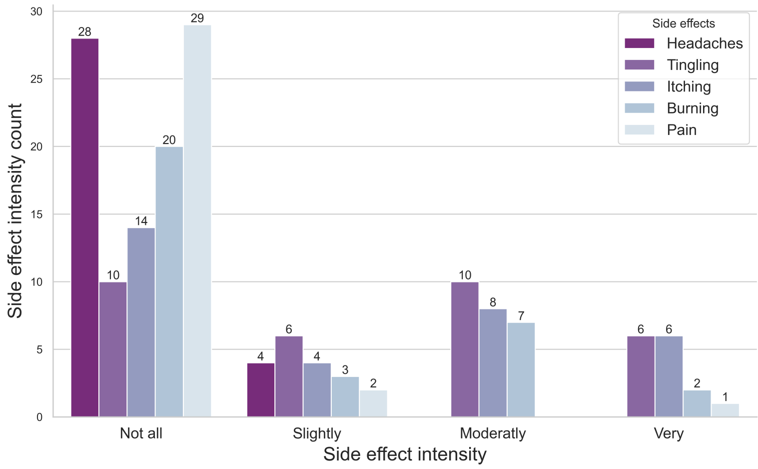
Table 4.3: Median intensity of the tDCS side effects

A Wilcoxon test for the side effects across active and sham for burning (p = .40), itching (p = .92), headache (p = .40), and pain (p = .12) showed no differences.

In Figure 4.12, we show the side effects intensity ratings for all our participants.



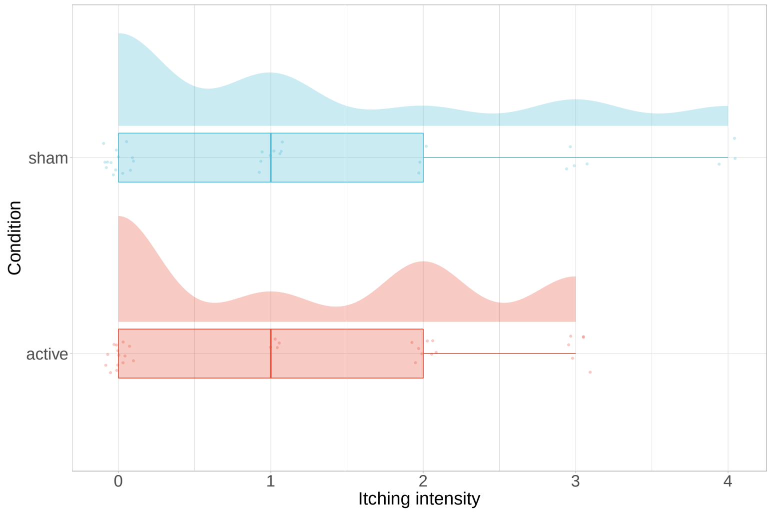
(a) Sham



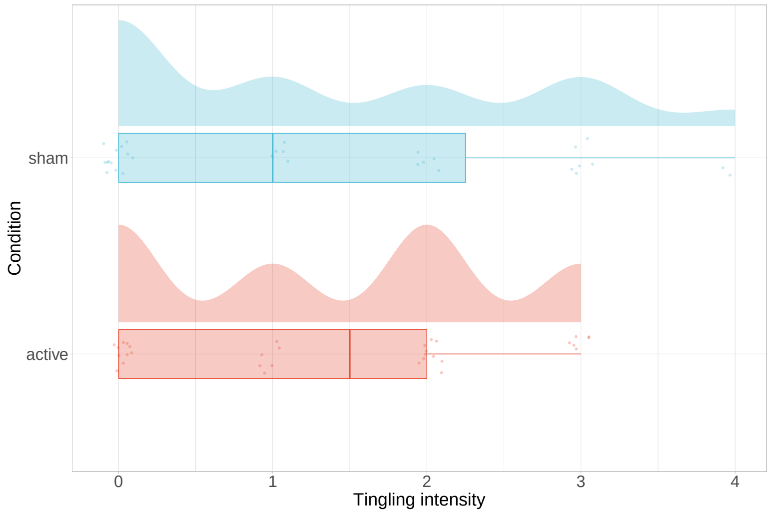
(b) Active

Figure 4.12: Side effects' intensity breakdown

Overall, we can see that across both conditions, the most common side effects are itching and tingling, they also have the highest number of participants that felt the strongest intensity. When looking at a breakdown of the intensity reporting for tingling and itching in Figure 4.13, we can see peaks for active for both side effects after the median which shows a group of people felt a higher intensity. As reported previously, this difference was not statistically significant.



(a) Itching intensity rating.



(b) Tingling intensity rating.

Figure 4.13: Itching and tingling intensity ratings.

4.3.1.6 Condition guess

After the second (final) session we asked our participants in the exit questionnaire if they could guess which session involved sham stimulation. We looked at the correct condition guess as the subjective belief of being under stimulation could moderate the effect of tDCS [287].

We find that 19 participants guessed correctly which session was sham and 13 were wrong in their guess. Slightly more than half of the participants correctly guessed

(N=19) if they were performing the task under sham or active stimulation. As part of our analysis, we looked at the potential side effects which could have helped the participants build their judgement to make their condition guess. Since the intensity ratings were not normally distributed, we used a non-parametric test. We used the two samples Wilcoxon rank sum test (also known as Mann-Whitney test) to investigate whether there was a statistically significant difference in the intensity rating of each side effect of the participants who correctly guessed with the rating intensity of the participants who were wrong in their guess.

Participants who correctly guessed their condition felt a stronger burning intensity (Mdn = 0.5,IQR = 2) compared to participants who poorly guessed (Mdn = 0,IQR = 0.75), this difference is statistically significant (z=-2.01, p=.02). We do not find any statistically significant difference for the rest of the side effects. In the table below (Table 4.4), we report the median intensity rating for each side effect.

Side effect type	**Correct condition guess**	**Incorrect condition guess**
Burning intensity	0.5* (2)	0 (0.75)
Itching intensity	1 (2)	0 (1.75)
Tingling intensity	1 (2.75)	1 (2)
Headache intensity	0 (0)	0 (0)
Pain intensity	0 (1)	0 (0)

Table 4.4: Median intensity ratings across the side effects for participants who correctly guessed the sham session and those who did not correctly guess. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

In Figure 4.14, we represent on the y-axis the participants who guessed correctly through a binary encoding: 1 represents correctly guessed, 0 represents did not guess correctly. In order to better understand how the burning intensity was experienced for the participants, we also show the breakdown across conditions. Participants who guessed correctly had a stronger burning intensity rating in active (Mdn = 1,IQR = 2) than sham (Mdn = 0,IQR = 1.5), this change however was not statistically significant (z=-0.35,p=.36).

Out of the 13 participants who did not guess correctly their conditions, we observe that overall they were less sensitive to burning. Participants who correctly guessed

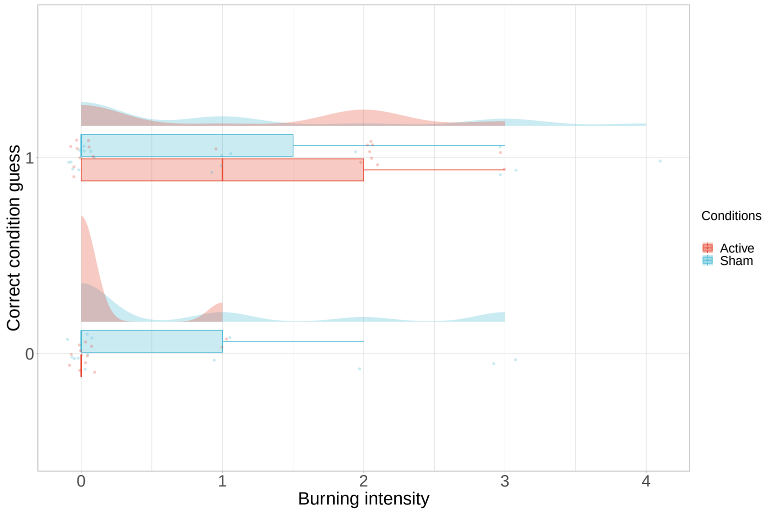


Figure 4.14: Raincloud representation of the burning intensity of the participants who guessed correctly and of the ones who did not guess correctly across active and sham

(N=19) were more sensitive to burning regardless of the condition, however we can observe that there is a right skeweness for these participants under the sham condition while it appears that under the active condition there is a more normally distributed burning intensity.

A further statistical analysis was performed on the side effects duration ratings. As per the last analysis, since our data is not normally distributed, we use the Wilcoxon rank sum test. We find that the burning duration rating of the participants who correctly guessed the sham session felt the burning for a longer period ($Mdn = 0.5$,$IQR = 2$) compared to the participants who did not guess correctly ($Mdn = 0$,$IQR = 0.75$), this change is statistically significant (z=-2.09,p=.02). In addition, we find a statistically significant higher itching duration (z=-1.83,p=.03) for the participants who guessed correctly ($Mdn = 1$,$IQR = 2$) compared to those who incorrectly guessed ($Mdn = 0$,$IQR = 2$). The rest of the side effects duration did not show any statistically significant change (Table 4.5).

Side effect type	**Correct condition guess**	**Incorrect condition guess**
Burning duration	0.5* (2)	0 (0.75)
Itching duration	1* (2)	0 (2)
Tingling duration	1 (2)	1 (1)
Headache duration	0 (0)	0 (0)
Pain duration	0 (0)	0 (0)

Table 4.5: Median duration ratings across the side effects for participants who correctly guessed the sham session and those who did not correctly guess. '*' indicates $p \leq .05$, '**' indicates $p < .01$, '***' indicates $p < .001$.

4.3.2 Pre-registered analysis

In this section, we present our pre-registered analysis focusing on dwell time. The dwell time is the time spent on the individual search results that the users' visited, excluding the time spent on any page of the search engine (excluded the home page of Moogle and its results pages).

For our analysis, we calculated the *total* dwell time, which is the total time spent on all the search results visited.

Our pre-registered hypothesis was: "There will be a larger reduction of total dwell time from baseline to the end of stimulation after 15min of 2 mA anodal tDCS on the left DLPFC compared to a sham protocol". As highlighted in the previous sections, this hypothesis was based on the close association of the left DLPFC with the executive functions and, in particular, working memory.

A dwell time change score (ΔDT) was calculated by subtracting the total dwell time in search task 1 from the total dwell time in search task 2 (active - base_active, sham - base_sham). We then performed a one-tailed, repeated measures t-test to assess the difference in this change between the active and sham tDCS conditions. The results from the active session (M=9280.13 ms, SD=109071.40 ms) and sham (M=-13472.47 ms, SD=99958.54 ms) indicate that anodal stimulation did not result in a reduction of the dwell time, $t(31) = 1.22$, $p = .23$. These results suggest that an anodal tDCS stimulation of the left DLPFC at a current of 2 mA for a duration of 15 minutes does not induce a reduction of the total dwell time of participants. We plotted the dwell

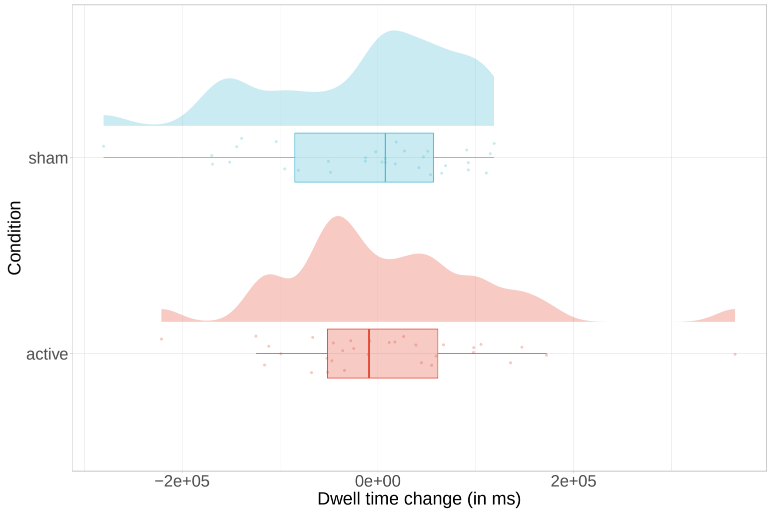


Figure 4.15: Dwell time change across tDCS conditions.

time change score in Figure 4.15. A negative change score represents a reduction in the dwell time.

In addition, we also compared the participants' total dwell time across conditions. We compared the active stimulation block (M=354244.3 ms, SD=58840.2 ms) to the block completed prior to active stimulation which we denoted as base active (M=344964.2 ms, SD=79132.19 ms), sham stimulation block (M=347234.8 ms, SD=72528.78 ms) to base sham (M=360707.2 ms, SD=69848.19 ms), and active to sham. We found that the total dwell time differences showed no change for any of the different combinations (Table 4.6).

	t(31)	**p-value**
active − base_active	0.48	0.63
sham − base_sham	-0.76	0.45
active − sham	-0.40	0.69

Table 4.6: Paired t-test across the different conditions and their control task

4.3.3 Exploratory analysis

Among the outcome measures we identified, we believed that each one could potentially be impacted by anodal stimulation of the prefrontal cortex. We believed the dwell time would be impacted because as participants stored more information in their working memory, they would be faster at processing and judging relevance of a web page. A similar reasoning can be applied to the SERP dwell time as participants should spend less time browsing the search results between their navigation as they should be able to recognise previously seen search results and filter out faster through identified useful keywords. Regarding queries, an improvement of the working memory would lead the participants in refining their questions as they collect new relevant keywords which would narrow the search path and therefore iterate more in the number of queries formulated and the queries might be different as they build their knowledge, they would also spend less time in writing their queries due to an improved vision of the search domain. Finally, regarding the websites they visited, the same reasoning as for the SERP applies; as participants build their knowledge they should be able to filter out the search results better and spend less time exploring the search results page (reduction in hovering) we would assume that searchers would not visit the same websites, and improve in filtering out results.

Therefore, we performed an exploratory analysis across the outcome measures extracted from the participants' log data during their exploratory search (subsection 4.2.10) and calculated their change score by subtracting the data collected during the base active tDCS condition from the data obtained during the active stimulation period (active - base_active), and by subtracting the base sham tDCS session data from the sham stimulation period data (sham - base_sham). For our exploratory analysis, we used Linear Mixed Models (LMMs) to identify whether there was a tDCS conditions effect, because LMMs allow us to integrate random effects (in this case participants) and they also allow us to identify potential tDCS moderating effects by integrating them as additional fixed effects.

In addition to the dwell time change (ΔDT), we analysed the following dependent variables: SERP dwell time (ΔSERP), number of queries change score (ΔQ), queries similarity change score (ΔQS) query formulation time change score (ΔQF), number of websites visited change score (ΔNWV), number of URLs hovered on change score (ΔNUHO), number of unique URLs hovered on change score (ΔNUUHO), number of actions change score (ΔNA).

We aimed to investigate whether tDCS had an impact on these outcome measures by adding participants as a random effect to control for the individual differences as it had been shown in previous studies that tDCS suffers from inter-subject variability [288]. Furthermore, we investigated the impact of three potential moderating factors of active anodal tDCS stimulation: feeling side effects (intensity and duration), correct condition guess, and BIS/BAS results.

We grouped our outcome measures into three groups: Dwell time and SERP dwell time, queries, interaction with websites. For each group, we report the random effects' variance and the results of our models.

Our analysis is structured as follows: We test for an effect of tDCS conditions by defining a model without any fixed effect and the participants as a random effect (Equation 4.1), we perform a likelihood-ratio test with a model that has tDCS conditions (active anodal and sham) as a fixed effect (Equation 4.2).

$$metric_x < -1 + (1|participants) \tag{4.1}$$

Subsequently, to test for the moderating effects, we use the latter model as our reference model (i.e model with tDCS conditions as fixed effect, see Equation 4.2). The results of our reference models are presented across the three groups of outcome measures (Dwell time and SERP dwell time, queries, and interactions with websites), we follow up by building extended models with the moderating factors (side effects, correct condition guess, and BIS/BAS scale) and compare the models' to the reference models using the likelihood-ratio tests. This test is used to compare two models, our reference model (with tDCS conditions, active anodal and sham, as fixed effect) that is lacking moderating factors with the model which has the additional fixed effects of interest (moderating factors). When we perform the test, we investigate whether the model with the moderating factors is significantly different from the one without [285].

We defined our LMMs with our outcome measures as our dependent variables. In our reference model, our dependent variables would be set as $Y = metric_x$, and the fixed effect would be the condition. The advantage of using LMMs is that we could specify the participants as a random effect to take into account for the variability across participants. Our reference model looks like this:

$$metric_x < -conditions + (1|participants) \tag{4.2}$$

Subsequently, we define our extended models which integrates each of our identified moderating factors:

$$metric_x < -conditions * side\ effects_{intensity_n} + (1|participants) \tag{4.3}$$

$$metric_x < -conditions * side\ effects_{duration_n} + (1|participants) \tag{4.4}$$

$$metric_x < -conditions * (\textstyle\sum_{n=0}^{4} side\ effects_{intensity_n} + side\ effects_{duration_n}) + (1|participants) \tag{4.5}$$

$$metric_x < -conditions * correct\_condition\_guess + (1|participants) + (1|topics) \tag{4.6}$$

$$metric_x < -conditions * (BIS\ score + BAS\ score) + (1|participants) + (1|topics) \tag{4.7}$$

Across the models defined above, the only difference compared to Equation 4.2 is the addition of the identified potential moderating factors as fixed effects. In Equation 4.5, we want to find out how our conditions effect varied across all the side effects with the intensity and duration combined, while in Equation 4.3 and Equation 4.4 we separate intensity and duration.

In the second model (Equation 4.6), we replace the side effects with our binary variable ‘correct condition guess’. Finally, we look at the impact of the BIS/BAS score (Equation 4.7).

In the models with moderating factors that we defined, we used the ‘*’ between the fixed effects (Equation 4.9a), this allows us to test the effect of the moderating factor as well as its interaction with the conditions (active/sham), in the case of side effects, it would be the same as writing:

$$conditions + side\_effect + conditions : side\_effect \tag{4.8}$$

If we found a significant difference with the reference model, we subsequently defined the same extended model but with the ‘+’ to test whether the interaction of the moderating factor lead to a statistically significant difference using likelihood-ratio test (Equation 4.9b).

$$conditions * side\_effect \tag{4.9a}$$

$$conditions + side\_effect \tag{4.9b}$$

4.3.3.1 Dwell time and SERP dwell time

We defined and built our LMMs for both the dwell time change (ΔDT) and the SERP dwell time change (ΔSERP) as our dependent variables.

In Table 4.7, we report the results of our random effects variance and standard deviation for both outcome measures in the reference models (models with the tDCS conditions as fixed effect): Across both outcome measures, we find that the LMMs we defined do not show a variance across participants. All of the variance is due to ‘residuals’, the residuals represent the random deviations from the predicted values which are not due to other random effects, in this case, to participants.

	ΔDT		**ΔSERP**	
	Variance	Std Deviation	Variance	Std Deviation
Participants	0.00	0.00	0.00	0.00
Residuals	10940	104.60	455.20	21.34

Table 4.7: Random effects variance and standard deviation across ΔDT and ΔSERP models without moderating factors

Dwell time change

A likelihood-ratio test between the model without any fixed effect and the one with conditions as a fixed effect showed that there was no difference between the two models and that therefore 2 mA of anodal tDCS stimulation did not result in a reduction of the dwell time between the anodal tDCS stimulation session and sham tDCS stimulation session, $\chi^2(1) = 0.78$, $p = .38$.

To test for the effect of the various side effects intensities, we performed a likelihood-ratio test between our reference model and the model with each side effect intensity (Equation 4.3). In addition, we perform a likelihood-ratio test between the reference model and the model with side effect duration (Equation 4.4) to test for the side effects duration. Across our five side effects intensity based models, our likelihood-ratio test showed the model which integrates tingling as a fixed effect was different from our reference model (i.e. model with tDCS conditions as the only fixed effect), $\chi^2(2) = 6.54$, $p = .04$.

However, there was no difference between the reference model and the rest of the extended models including the different intensity ratings for burning, itching, headache, and pain. In addition to testing for the intensity ratings of the side effects, we test for the duration of the side effects. None of the side effect durations were different to the reference model (Table 4.8).

	χ^2	df	p
Burning intensity	0.06	1	.97
Itching intensity	5.89	2	.05
Tingling intensity	6.54	2	.04 *
Headache intensity	0.45	2	.80
Pain intensity	0.05	2	.97
Burning duration	0.85	2	.65
Itching duration	5.47	2	.06
Tingling duration	3.18	2	.20
Headache duration	0.29	2	.87
Pain duration	0.24	2	.89
$side\ effects_{intensity} + side\ effects_{duration}$	30.65	19	.04
Correct condition guess	0.03	2	.98
$BIS_{score} + BAS_{score}$	2.75	4	.60

Table 4.8: LMMs results for the ΔDT with the different side effects intensity ratings as fixed effects. '*' indicates $p \leq .05$, '**' indicates $p < .01$, '***' indicates $p < .001$.

To test whether the interaction of tingling with conditions was statistically significant, we used a likelihood-ratio test: We did not find a difference between the model where tingling has an interaction with the conditions and the one without the interaction, $\chi^2(1) = 2.50$, $p = .11$. This indicates that tingling intensity did not moderate the effect of tDCS, and that when participants felt tingling, across both conditions, they changed their behaviour, ($\hat{\beta}$ = -39.85, SE = 16.08, t =-2.479). We represent the predictions of our model of the dwell time change in Figure 4.16, we found that as tingling intensity increases, the dwell time change decreases: This means that the dwell time in the active and sham conditions was lower than during base active (in respect to the active condition) and base sham (in respect to the sham condition). In other words, participants spent less time reading the web pages accessed (i.e decrease of dwell time) when there is a sensation of tingling.

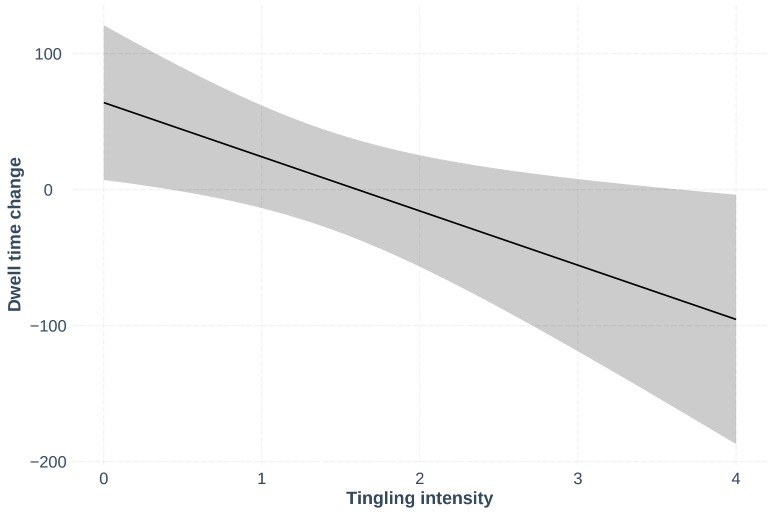


Figure 4.16: LMM prediction of dwell time change (in seconds) in respect to tingling intensity with tDCS conditions and tingling intensity as fixed effects. In grey is the 95% confidence interval.

To evaluate whether the combination of all the side effects intensities and duration had a moderating effect, we integrated all of the intensity ratings and duration ratings in one model and performed a likelihood-ratio test with the reference model. We find that our extended model with all the side effects has a significant effect, $\chi^2(19) = 30.65$, $p = .04$. In addition, we find that the interaction of the intensity and duration ratings with the conditions has a significant effect and therefore that there is an inter-dependent relationship between the combination of side effects (duration and intensity) and the tDCS conditions, $\chi^2(9) = 17.90$, $p = .04$. In Table 4.9, we find that the participants, in our improved model, explain 26% of the variance of our model: This may be related to the inter-subject variability of response to tDCS [288, 289].

	ΔDT		**ΔSERP**	
	Variance	Std Deviation	Variance	Std Deviation
Participants	2133	46.19	0.00	0.00
Residuals	5838	76.41	490.70	22.15

Table 4.9: Random effects variance and standard deviation across ΔDT and ΔSERP models with the side effects as additional fixed effects

A likelihood-ratio test across our reference model and the model extended with the correct condition guess shows no difference across the two models, $\chi^2(2) = 0.03$, $p = .98$. A similar analysis was performed on an extended model with the BIS and BAS scores, as fixed effect each, we did not find any difference with the reference model, $\chi^2(4) = 2.75$, $p = .60$.

SERP dwell time

In Figure 4.17, we represent the SERP dwell time change across conditions, we observe that the median is higher for the active session ($Mdn = 6.402$, $IQR = 28.99$) than the sham session ($Mdn = 3.03$,$IQR = 21.82$) with the tail of the distribution on the left side for active, this indicates that more participants had a lower SERP dwell time change in active compared to sham and that in the active stimulation participants had a longer SERP dwell time compared to base active task. In Figure 4.15, we have seen that participants had a decrease in the dwell time from baseline to the stimulation in the active session compared to the sham session where we did not have a decrease.

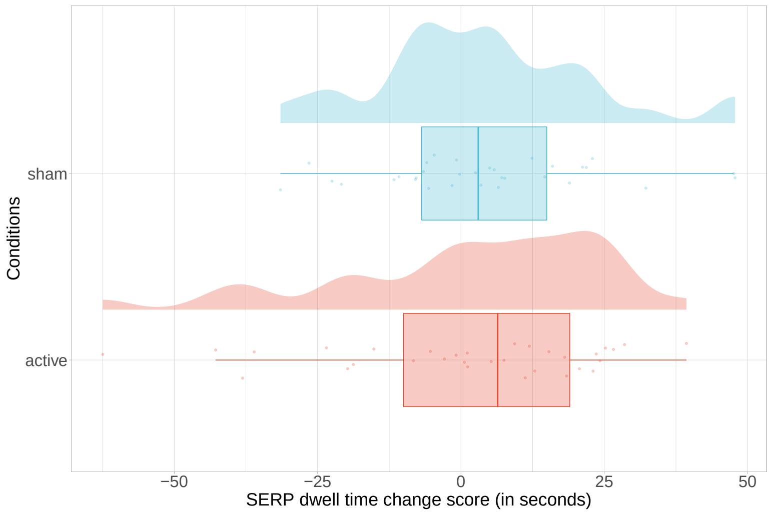


Figure 4.17: ΔSERP across conditions.

A likelihood-ratio test between our model without any fixed effect and the model with the conditions as a fixed effect has shown that there is no difference between the two models and that therefore, conditions did not have an effect in moderating the

SERP dwell time, $\chi^2(1) = 0.22$, $p = 0.64$.

We built our baseline LMM for the SERP dwell time, in Table 4.7 we found that similar to the dwell time model, there is no variance for the participants.

The likelihood-ratio test between our reference model and the models extended with the various side effects intensity ratings did not show any difference (Table 4.10).

	χ^2	df	p
Burning intensity	3.22	2	.20
Itching intensity	2.71	2	.26
Tingling intensity	4.20	2	.12
Headache intensity	0.48	2	.79
Pain intensity	4.42	2	.11
Burning duration	4.97	2	.08
Itching duration	1.23	2	.54
Tingling duration	2.17	2	.34
Headache duration	0.38	2	.82
Pain duration	3.96	2	.14
$side\ effects_{intensity} + side\ effects_{duration}$	17.95	19	.53
Correct condition guess	3.89	2	.14
$BIS_{score} + BAS_{score}$	0.37	4	.98

Table 4.10: LMMs results for the ΔSERP with the different side effects intensity ratings as fixed effects. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

In line with the findings with the models extended with side effects intensity ratings, we did not find any difference between the models with the side effects duration and the reference model (Table 4.10). When performing the likelihood-ratio test between the model with all the side effects intensity ratings and duration, we did not find any difference with the reference model, $\chi^2(19) = 17.95$, $p = .53$. The analysis across all the side effects indicates that none of the side effects had a moderating effect on the SERP dwell time change.

Finally, neither the correct condition guess had a moderating effect on the SERP dwell time, $\chi^2(2) = 3.89$, $p = .14$, nor the BIS and BAS score, $\chi^2(4) = 0.37$, $p = .98$.

4.3.3.2 Queries

For the 'queries' group of outcome measures, we apply the same approach as for the dwell time outcome measures and build a similar LMM (Equation 4.2) with conditions as fixed effects and participants as the random effect for the reference model.

In Table 4.11, we look at the users' random effect variance across the 'queries' outcome measures: The number of queries change (ΔQ) LMM did not show any variance for participants and only a small one for the query formulation time change (ΔQF). Finally, for the query similarity change (ΔQS), participants account for 46% of the variance, this indicates an inter-participants variability in the way queries were formulated across search tasks.

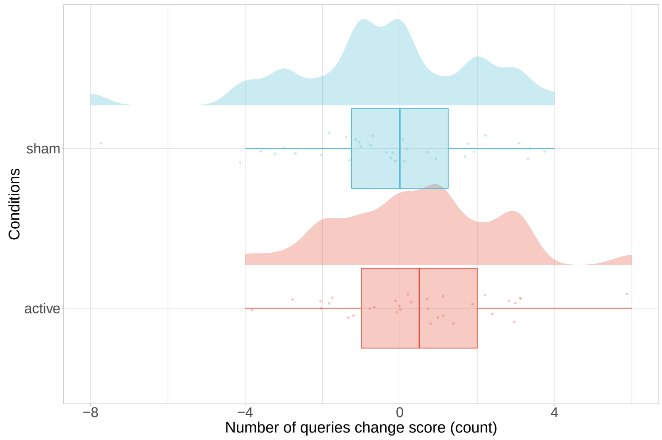
	ΔQ		**ΔQS**		**ΔQF**	
	Variance	Std Deviation	Variance	Std Deviation	Variance	Std Deviation
Participants	0.00	0.00	0.11	0.34	0.00	0.05
Residuals	5.35	2.313	0.13	0.36	0.04	0.19

Table 4.11: Random effects variance and standard deviation across ΔQ, ΔQS, and ΔQF models without moderating factors

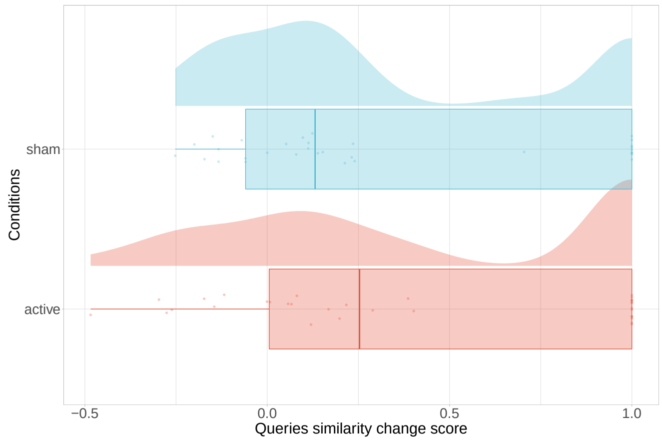
We plot the number of queries change, queries similarity change, and query formulation time change. Figure 4.18a indicates that for the number of queries change in the active session ($Mdn = 0.5$, $IQR = 3$) less queries were formulated in the base active condition than in the active condition. While the number of queries change in the sham session ($Mdn = 0$, $IQR = 2.5$) shows about the same number of queries were formulated across both sham and base sham.

Regarding the query similarity change, in Figure 4.18b we see a similar trend across the active and sham sessions. However, it appears that queries were more similar in query similarity change in the active session ($Mdn = 0.253$, $IQR = 0.995$) than in the query similarity change in the sham session ($Mdn = 0.131$, $IQR = 1.059$). However, across both number of queries change and query similarity change, the change is small.

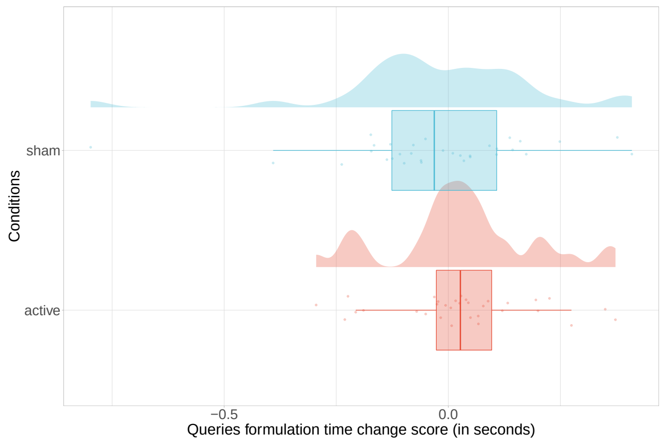
Finally, we observe that for query formulation time change, the query formulation time change was longer in the active session ($Mdn = 2.318$, $IQR = 20.964$) for the active condition than the base and the opposite effect in the sham session change score (Mdn = -4.805, IQR = 16.344) where base sham had a longer formulation time.



(a) ΔQ across both the sham and active session.



(b) ΔQS across both the sham and active session.



(c) ΔQF across both the sham and active session.

Figure 4.18: Queries related outcome measures change score across the sham and active session.

Number of queries

We perform a likelihood-ratio test across the reference model and the model without any fixed effect to test for tDCS stimulation condition effect, we do not find any difference between the two models, $\chi^2(1) = 2.32$, $p = .13$. In addition, a likelihood-ratio test across the side effects intensity ratings with the reference model has shown that there is no moderating effects due to the side effects intensity ratings (Table 4.12).

	χ^2	df	p
Burning intensity	1.44	2	.49
Itching intensity	2.57	2	.28
Tingling intensity	2.05	2	.36
Headache intensity	.45	2	.80
Pain intensity	5.57	2	.06
Burning duration	3.55	2	.17
Itching duration	.60	2	.74
Tingling duration	1.17	2	.56
Headache duration	.43	2	.81
Pain duration	6.17	2	.05*
$side\ effects_{intensity} + side\ effects_{duration}$	19.91	19	.40
Correct condition guess	.70	2	.70
$BIS_{score} + BAS_{score}$	3.17	4	.53

Table 4.12: LMMs results for the ΔQ with the different moderating effects as fixed effects. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

In Table 4.12, we find that our likelihood-ratio test shows a difference between our reference model and the model extended with the pain duration, $\chi^2(2) = 6.17$, $p = .05$.

While the pain intensity rating did not have any difference with the reference model, it is the one with the smallest p-value across the intensity ratings. These two findings indicate that the feeling of pain was likely involved in moderating the participants behaviour. In Figure 4.19, we observe that as pain duration increases, the number of queries change increases which indicates that as participants were under more pain, they formulated more queries under the tDCS active and sham stimulation. The model summary corroborates this observation, as the pain duration increases by one, the number of queries change increases by 0.68 ($\hat{\beta} = 0.68$, $SE = 0.53$, $t = 1.29$).

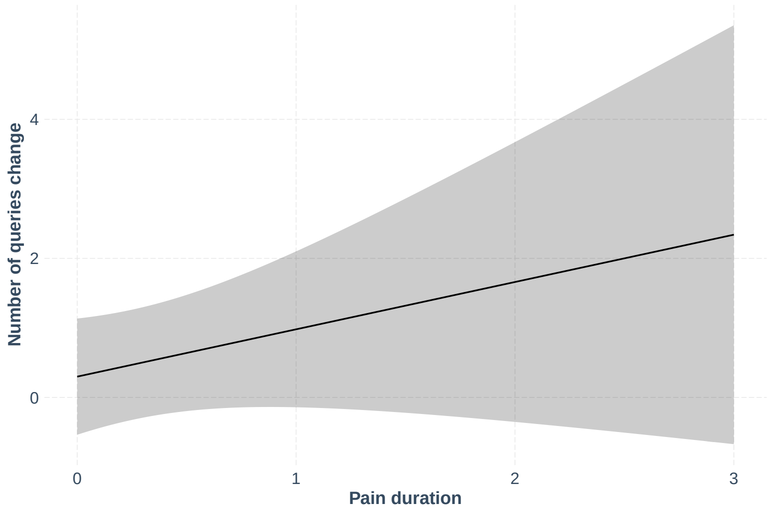


Figure 4.19: LMM prediction of number of queries change in respect to pain duration with tDCS conditions and tingling pain duration as fixed effects. In grey is the 95% confidence interval.

Subsequently, we tested whether the interaction between pain duration and conditions is significant. The likelihood-ratio test across the model with the pain duration interacting with tDCS conditions and the model without the interaction showed a difference between the two models, $\chi^2(1) = 5.88$, $p = .02$: This indicates that there is a significant interaction between pain duration and conditions.

Our model with all our side effects did not have any difference with the model without the side effects, $\chi^2(19) = 19.91$, $p = .40$ (Table 4.12). No difference across the models with correct condition guess, BIS and BAS score and the reference model was found indicating no moderating effect.

Queries similarity

When testing for a tDCS conditions effect, our likelihood-ratio test across the model without the conditions and the model with the conditions indicated no difference across the models, $\chi^2(1) = 1.05$, $p = .30$. Further likelihood-ratio tests were performed to identify potential moderating effects for queries similarity, we did not find any difference across the various extended models defined and the reference model with the tDCS conditions (Table 4.13).

	χ^2	df	p
Burning intensity	1.73	2	.42
Itching intensity	1.96	2	.38
Tingling intensity	.52	2	.77
Headache intensity	.11	2	.94
Pain intensity	.63	2	.73
Burning duration	3.47	2	.18
Itching duration	1.01	2	.60
Tingling duration	2.52	2	.28
Headache duration	.43	2	.81
Pain duration	2.76	2	.25
$side\ effects_{intensity} + side\ effects_{duration}$	18.04	19	.52
Correct condition guess	.10	2	.95
$BIS_{score} + BAS_{score}$	5.17	4	.27

Table 4.13: LMMs results for the ΔQS with the different moderating effects as fixed effects. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

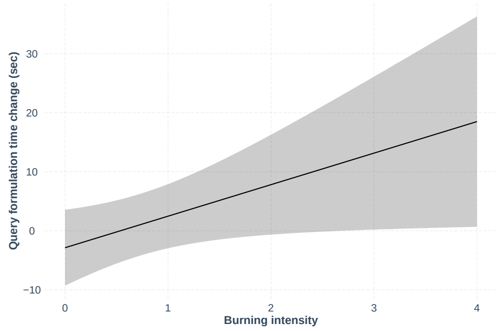
Queries formulation time

The likelihood-ratio test between the reference model with the tDCS conditions as fixed effects and the model without the tDCS conditions indicate no difference between the two models, $\chi^2(1) = 1.42$, $p = .07$. This indicates that there is no conditions effect on the queries formulation time change.

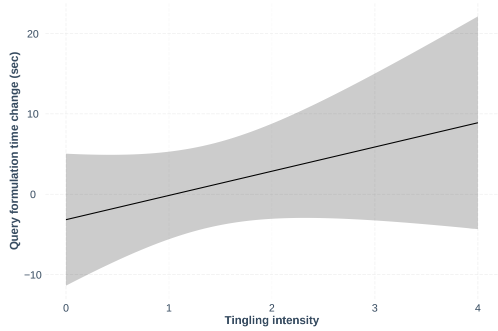
	χ^2	df	p
Burning intensity	8.01	2	.02*
Itching intensity	1.51	2	.47
Tingling intensity	9.69	2	.007**
Headache intensity	3.60	2	.16
Pain intensity	8.33	2	.01*
Burning duration	9.02	2	.01*
Itching duration	0.82	2	.67
Tingling duration	3.81	2	.15
Headache duration	3.44	2	.17
Pain duration	7.88	2	.02*
$side\ effects_{intensity} + side\ effects_{duration}$	36.71	19	.008**
Correct condition guess	0.37	2	.83
$BIS_{score} + BAS_{score}$	5.56	4	.23

Table 4.14: LMMs results for the ΔQF with the different moderating effects as fixed effects. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

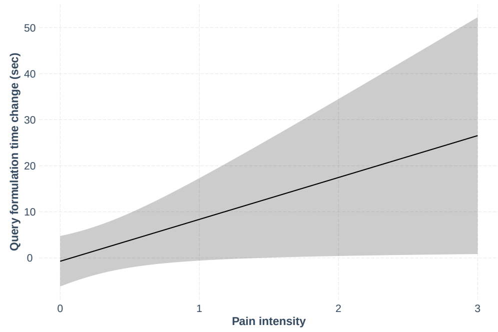
In Table 4.14, we find that for the side effects intensity ratings, three side effects had a moderating effect on the query formulation time: Burning, tingling, and pain. The model with the burning intensity rating was different to our reference model, $\chi^2(2) = 8.01$, $p = .02$. The summary of the model with the burning intensity indicated that as burning increases by one unit, the query formulation time change increased by 5.34 seconds ($\hat{\beta} = 5.34$, $SE = 2.59$, $t = 2.06$). The model with the tingling intensity was also different from the reference model, $\chi^2(2) = 9.69$, $p = .007$, and had a similar impact to burning whereas the tingling intensity increased by one unit, the query formulation time change increased by 3 seconds ($\hat{\beta} = 3.02$, $SE = 2.32$, $t = 1.30$). Finally, the same effect was found for pain intensity, the model that integrated pain intensity was different from the reference model ($\chi^2(2) = 8.33$, $p = .01$), and as the pain intensity increased by one unit, the query formulation time change increased by 9 seconds ($\hat{\beta} = 9.09$, $SE = 4.47$, $t = 2.03$). Across these three side effects intensity ratings, we plot the models' predictions in Figure 4.20.



(a) Burning intensity rating



(b) Tingling intensity rating

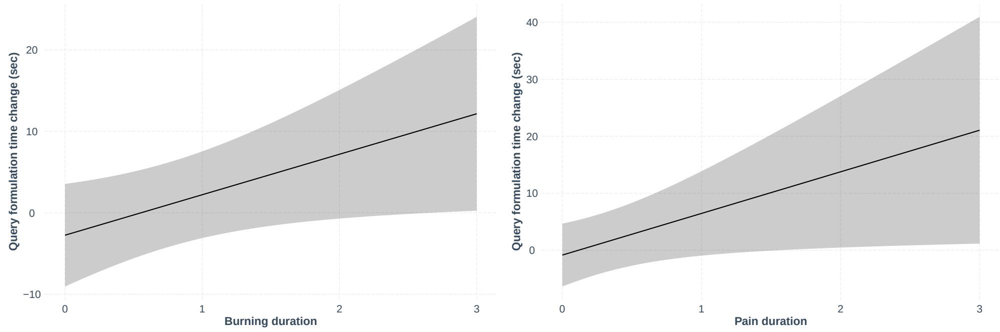


(c) Pain intensity rating

Figure 4.20: Query formulation time prediction given burning, tingling, and pain intensity ratings. In grey is the 95% confidence interval.

Looking at the side effects duration, we have found that burning duration model ($\chi^2(2) = 9.02$, $p = .01$) and the pain duration model ($\chi^2(2) = 7.88$, $p = .02$) were different from our reference model. We find that as a burning duration increases by one unit, the query formulation time change increases by 4.97 seconds ($\hat{\beta} = 4.97$, $SE = 2.38$, $t = 2.09$) and as pain duration increases by one unit, the query formulation time changes by 7.30 seconds ($\hat{\beta} = 7.30$, $SE = 3.49$, $t = 2.09$). In Figure 4.21, we can visually observe our models' predictions across the two side effects.

These findings across the intensity ratings and duration show that as intensity ratings and duration increased, the query formulation time change increased as well and therefore, that under active and sham tDCS stimulation participants took more time to formulate their queries compared to the base conditions when they were not stimulated. This is counter-intuitive as we would have expected that participants with improved working memory would have an improved decision making which would allow them to decide faster on the next step they will take in the search process. However, this could mean that when participants felt these side effect, they have put in more effort in formulating their queries as they would believe they were stimulated.



(a) Burning duration rating

(b) Pain duration

Figure 4.21: Query formulation time prediction given burning, and pain duration. In grey is the 95% confidence interval.

Further likelihood-ratio tests were performed to test for the interaction of conditions and the various side effects intensity and duration which had an effect on the query formulation time. We find a difference across the models with the interaction and the models without, we report the results in Table 4.15.

	χ^2	df	p
Burning intensity:conditions	7.97	1	.004**
Tingling intensity:conditions	7.98	1	.005**
Pain intensity:conditions	8.04	1	.005**
Burning duration:conditions	8.93	1	.003**
Pain duration:conditions	7.84	1	.005**

Table 4.15: LMMs results for the ΔQF testing for the interaction of tDCS conditions and the relevant side effects intensity ratings and duration. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

Finally, we also found a difference between the reference model and the model with all the side effects intensity ratings and duration, $\chi^2(19) = 36.718$, $p = .008$. In addition, we found that there is an interaction of the tDCS conditions and the various side effects combined, $\chi^2(9) = 27.534$, $p = .001$.

In Table 4.16, we represent the random effects variance for the model with all the side effects with the interactions with the tDCS conditions, and we did not find any participants variance, so it appears that there was no inter-participants variability for the query formulation time.

	Variance	Std Deviation
Participants	0.00	0.00
Residuals	194.7	13.95

Table 4.16: Random effects variance and standard deviation for ΔQF models with the side effects intensity ratings and duration as moderating factor

Additional likelihood-ratio test did not show any difference between the models with the correct condition guess, BIS and BAS score, and the reference model.

4.3.3.3 Interaction with websites

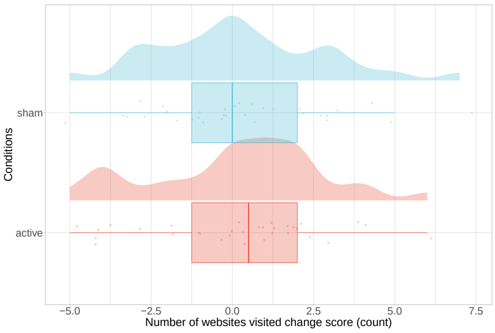
As highlighted in the introduction of this section, we have identified a set of related outcome measures related to the web pages which were accessible through the search results pages. In Table 4.17, we found that there was no variance in our LMMs that is due to the participants.

	ΔNWV		**ΔNUWV**		**ΔNUHO**		**ΔNUUHO**			
	Variance	Std Deviation	Variance	Std Deviation	Variance	Std Deviation	Variance	Std Deviation	Variance	Std Deviation
Participants	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Residuals	0.06	0.25	6.88	2.62	5.08	2.25	912.5	30.21	46.15	6.794

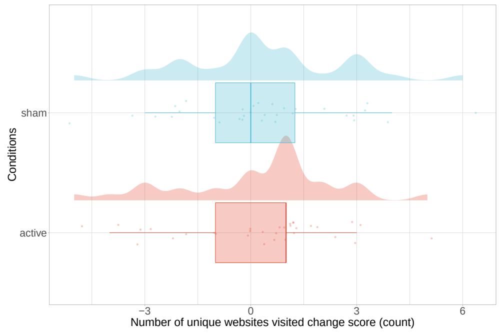
Table 4.17: Random effects variance and standard deviation across ΔNWV, ΔNUWV, ΔNUHO, and ΔNUUHO models without moderating factors

Regarding ΔNWV and ΔNUWV, we observe a similar pattern across the outcome measures. A right skewness for the change score of the sham session and a left skewness for the active session: This would indicate that more websites were visited under the active condition than the base active condition and that more websites were visited under the base sham condition compared to the sham condition.

Meanwhile, regarding the hovering related outcome measures (ΔNUHO and ΔNUUHO), we find similar trend for the active session with the right skewness which would indicate

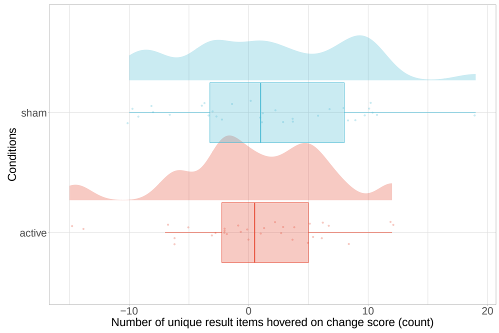


(a) ΔNWV



(b) ΔNUWV

(c) ΔNUHO



(d) ΔNUUHO

Figure 4.22: Web results related outcome measures change score across the active and sham session.

more links were hovered on under active than base active. The same can be observed for the sham session but without any skewness which would indicate we had a fairly well distributed and varied interaction during hovering across sham and base sham.

We build LMMs across the various outcome measures to verify our observations.

Number of websites visited

A likelihood-ratio test to verify whether conditions had an effect on the number of websites visited did not find any difference, $\chi^2(1) = .04$, $p = .85$. A similar test was performed on the number of unique websites visited and similarly, we did not find any difference across baseline and the model with the tDCS conditions, $\chi^2(1) = .08$, $p = .78$.

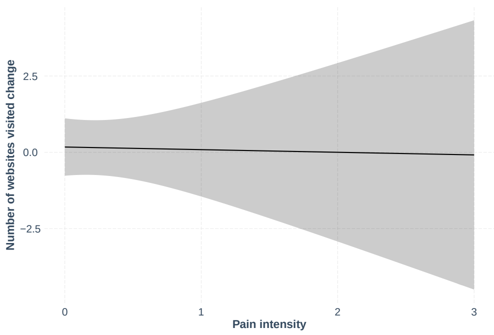
	ΔNWV			**ΔNUWV**		
	χ^2	df	p	χ^2	df	p
Burning intensity	4.31	2	.12	5.53	2	.06
Itching intensity	1.38	2	.50	1.71	2	.43
Tingling intensity	5.22	2	.07	6.03	2	.05*
Headache intensity	5.09	2	.08	4.26	2	0.12
Pain intensity	6.69	2	.04*	8.14	2	.02*
Burning duration	5.05	2	.08	7.55	2	.02*
Itching duration	.18	2	.91	.08	2	.96
Tingling duration	2.54	2	.28	2.74	2	.25
Headache duration	4.81	2	.09	4.27	2	.12
Pain duration	7.91	2	.02*	9.47	2	.008**
$side\ effects_{intensity} + side\ effects_{duration}$	22.65	19	.25	24.89	19	.16
Correct condition guess	2.00	2	.37	1.78	2	.41
$BIS_{score} + BAS_{score}$	.07	4	1.00	1.29	4	.86

Table 4.18: LMMs results for the ΔNWV and ΔNUWV with the different moderating effects as fixed effects. '*' indicates p $\leq$ .05 , '**' indicates p < .01, '***' indicates p < .001.

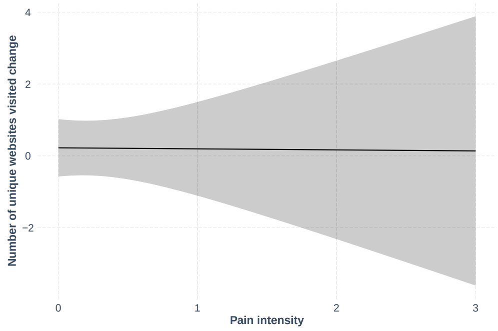
In Table 4.18, we find that across both the number of websites visited and the number of unique websites visited, the effect of the side effects is fairly consistent for both. We have a pain intensity rating effect, and pain duration effect for both outcome measures. The only side effects which were statistically significant only to the number of unique websites visited are tingling intensity, $\chi^2(2) = 6.03$, $p = .05$, and burning duration, $\chi^2(2) = 7.55$, $p = .02$. However, these still show a fairly small p-value for the number of websites visited.

In Figure 4.23, we see a similar pattern for both pain intensity rating and pain duration in respect to our web pages visited outcome measures. It appears to be that as pain duration ratings increased, the participants had an increase in the number of websites visited and unique websites visited, i.e as pain duration increases, participants visited more websites in active and sham tDCS stimulation (Table 4.18). While it appears to be that as pain increases, participants had as many websites visited across active and sham tDCS stimulation, and their base conditions.

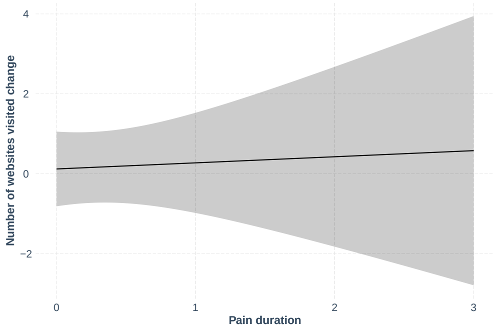
The model summary indicated that as pain increases by one unit, the number of websites visited decreased ($\hat{\beta}$ = -0.09, SE = 0.77, t = -0.11), and the number of unique websites decreased as well ($\hat{\beta}$ = -0.03, SE = 0.65, t = -0.04). While as pain duration increases by one unit, the number of websites visited increased ($\hat{\beta}$ = 0.15, SE = 0.59, t = 0.26)), and similarly the number of unique websites visited increased ($\hat{\beta}$ = 0.18, SE = 0.50, t = 0.36).



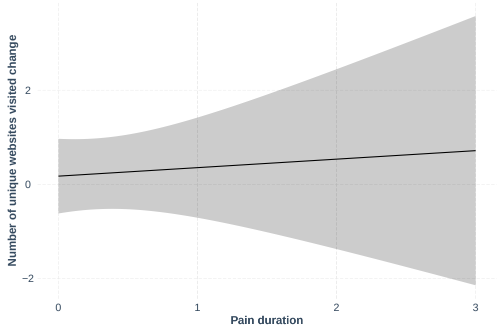
(a) Pain intensity for number of websites visited



(b) Pain intensity for number of unique websites visited



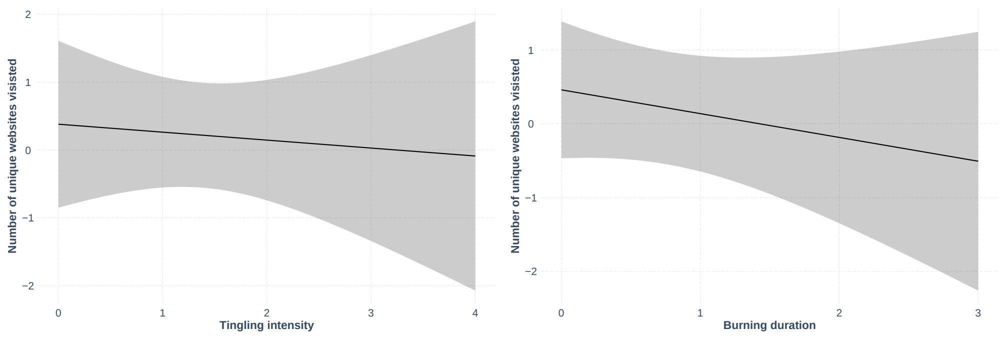
(c) Pain duration for number of websites visited



(d) Pain duration for number of unique websites visited

Figure 4.23: Number of websites visited and number of unique websites visited prediction given pain intensity and pain duration. In grey is the 95% confidence interval.

In Figure 4.24, we can see that across both tingling intensity and burning duration,



(a) Tingling intensity

(b) Burning duration

Figure 4.24: Number of unique websites visited prediction given tingling intensity, and burning duration. In grey is the 95% confidence interval.

these two side effects have a negative relationship with the number of unique websites visited. As tingling increased by one unit, the number of unique websites visited decreased ($\hat{\beta}$ = -0.12, SE = 0.35, t = -0.34). While as burning duration increased by one unit, the number of unique websites visited decreased ($\hat{\beta}$ = -0.32, SE = 0.35, t = -0.92).

An increase in the change across the two outcome measures means that participants visited more websites when active and sham stimulation. A decrease would represent a larger number of websites visited under base active and base sham. Therefore, we can say that when pain intensity ratings, tingling intensity ratings, and burning duration ratings were higher, participants' reduced their browsing, it could be interpreted that it negatively impacted their experience. While the pain lasting for longer only lead participants to browse more.

We observe in Table 4.19 and Table 4.20, when testing for these side effects and their interaction with the conditions, we find that pain duration moderates the tDCS condition across both the number of websites visited and the number of unique websites visited. While the rest of the side effects had their impact on the participants regardless of the condition, in this case a decrease of the number of websites visited when under tDCS stimulation.

	χ^2	df	p
Pain intensity:conditions	1.90	1	0.17
Pain duration:conditions	4.88	1	.03**

Table 4.19: LMMs results for the ΔNWV testing for the interaction of tDCS conditions and the pain intensity and duration ratings. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

	χ^2	df	p
Tingling intensity:conditions	1.74	1	.19
Pain intensity:conditions	2.56	1	.11
Pain duration:conditions	6.09	1	.01
Burning duration:conditions	2.26	1	.13

Table 4.20: LMMs results for the ΔNUWV testing for the interaction of tDCS conditions and the relevant side effects intensity and duration ratings. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

Across the rest of the potential moderating factors, we find no difference with our reference model as presented in Table 4.18

Number of search results hovered on

A likelihood-ratio test across the model with the tDCS conditions and a model without indicates that the tDCS conditions did not have an effect on the number of search results hovered on, $\chi^2(1) = .44$, $p = .51$, nor the number of unique search results hovered on, $\chi^2(1) = .48$, $p = .49$.

In addition, further likelihood-ratio tests across the reference model with the tDCS conditions and the extended models for both outcome measures, number of search results hovered on and number of unique search results hovered on, has shown that no moderating effect happened from the side effects, correct condition guess, nor the BIS/BAS scores (Table 4.21).

	ΔNUHO			ΔNUUHO		
	χ^2	df	p	χ^2	df	p
Burning intensity	1.43	2	.49	2.89	2	.24
Itching intensity	2.01	2	.37	3.25	2	.19
Tingling intensity	1.84	2	.40	4.93	2	.08
Headache intensity	.61	2	.74	1.47	2	0.48
Pain intensity	4.52	2	.10	5.44	2	.07
Burning duration	2.38	2	.30	4.81	2	.09
Itching duration	.24	2	.89	.70	2	.70
Tingling duration	2.38	2	.30	3.97	2	.14
Headache duration	.60	2	.74	1.09	2	.58
Pain duration	4.55	2	.10	5.66	2	.06
$side\ effects_{intensity} + side\ effects_{duration}$	16.58	19	.62	25.16	19	.16
Correct condition guess	5.05	2	.08	1.49	2	.48
$BIS_{score} + BAS_{score}$	.43	4	.98	2.58	4	.63

Table 4.21: LMMs results for the ΔNUHO and ΔNUUHO with the different moderating effects as fixed effects. '*' indicates p $\leq$ .05 , '**' indicates p $<$ .01, '***' indicates p $<$ .001.

4.4 Discussion & conclusion

The aim of this work was to use tDCS to improve the individuals' search experience in the context of an IR task, hence leading to overcoming one trigger of IO: Brain ability and cognition [9].

In the IR literature, both the DLPFC and working memory were shown to play an active role in the search tasks [124]. In addition, previous work in cognitive neuroscience has shown the efficacy of tDCS stimulation in modulating the activity of the left DLPFC and its impact on various executive functions such as working memory [232], attention [274], language comprehension [250], decision-making [269], and solution recognition [290]. The rationale behind our work was that users involved in exploratory search tasks would be able to retain more information in their working memory, in addition due to the overlap of attention and decision-making with working memory and their strong association with the prefrontal cortex [251, 269], we hypothesised that this would lead our searchers to better filter out previously seen information and identify relevant information faster. Filtering information, and solution recognition would fall under what is described as relevance judgment in IR, we have seen that one implicit feedback method to relevance is the dwell time (i.e time spent on a web page excluding the search engine results page). Hence, we believed that tDCS stimulation would lead to a reduction of the dwell time.

We found that 15 min anodal stimulation of tDCS at a current of 2 mA on the left DLPFC had no effect on the dwell time (RQ 1). In addition, we performed an exploratory analysis using LMMs on three types of outcome measures which were derived from our log data: Dwell time and SERP dwell time, queries, and websites visited. In this exploratory analysis, we found that tDCS stimulation did not lead to any change in behaviour across the 3 types of outcome measures (RQ 2). Rather, we found that the feeling of various side effects led to a change of behaviour: We found that headaches moderated the change of the dwell time, where the more intense the headaches our participants felt, the more their dwell time decreased, regardless of the condition (active/sham). One explanation could be that our participants could not focus due to the headaches, or alternatively that the perception of the pain led them to change their behaviour as they believed that they were getting stimulated (i.e. a placebo effect). The feeling of pain over a continuous period of time was also positively correlated with an increase in the number of queries under tDCS stimulation, we found that this increase varied based on the tDCS conditions. Similarly, to the increase of number of queries, we found an increase in the query formulation time; the latter is positively correlated with stronger burning, tingling, and pain intensity. It also positively correlated with the longer burning and pain durations. We found that all of these side effects moderated the way that participants reacted across the active and sham conditions. For the websites visited related outcome measures, we found a decrease in the number of websites visited when participants felt pain, tingling, and burning at stronger intensities, while pain over time resulted in more websites visited under tDCS stimulation.

Finally, when looking at which of these side effects could have contributed to correctly guessing which tDCS condition was active and which was sham, we found that intensity ratings of burning, and itching duration ratings were larger for participants who correctly identified the tDCS conditions that they had experienced. We also found that over half of our participants were able to correctly guess their condition. These findings highlight the potential risks of the tDCS side effects in distracting participants from their task but also potentially leading to placebo effects in their behaviour.

While brain stimulation and neuromodulation techniques could provide positive outcomes for individuals who suffer neurological disorders, there are also various ethical concerns that ensue such as the potential negative use by private entities into manipulating labour but also creating poor inhibitory behaviour to reach profit. In this case, we would be impacting the free will of individuals and these questions fall under the research of neuroethics [291]. In the case of online search task, we have to consider how neuromodulation might be abused to accentuate inequalities, e.g students who have access to this technology could perform better in their assignments compared to those who do not. While our findings were not conclusive, these are important ethical concerns to be taken into account as brain stimulation and neuromodulation become more popular and effective. A first step which would lead to better scoping of the ethical issues associated with neurostimulation would be to better understand its impact on the brain. As we have seen, the variations in the outcomes of tDCS stimulation indicate that we do not yet have a good grasp of the technology.

There has been a lot of work done with brain stimulation, which shows its impact on performance in various brain regions, but this is the first to apply it in the context of IR. We hypothesised that anodal stimulation of 15 minutes would reduce the total dwell time of users, however this hypothesis was not supported. We believe this work opens a new window of research to the IR community. In future work, the various changes in behaviour related to side effects should be paid closer attention as this could lead to worse performance from the subjects. The use of brain stimulation methods with lower side effects, or the use of topical anaesthesia to reduce sensory side effects, may be beneficial in achieving more meaningful results. A better understanding of tDCS and how its parameters work is also necessary to potentially overcome the high variability in the studies using tDCS [288].

4.4.1 Limitations

One reason which could have led to our null results could be due to parameter settings of tDCS that were selected. Various parameters are involved in tDCS stimulation, the choice of brain region but also the size of the electrodes, current intensity, current density, and the duration of the stimulation. In addition, there are two broad types of designs when working with tDCS, one is online stimulation and one is offline stimulation. The use of online stimulation applies to our study, as the participants were completing the task as they were receiving stimulation. However, with such design as we have found there is the risk of the tDCS side effects being a distraction to our participants: This would lead to behavioural changes which could interfere with the results. This problem could be overcome by the use of an offline design (participants performing the task after stimulation): Participants have a pre-stimulation task and post-stimulation task, during the stimulation participants can either complete a filler task or remain inactive. In our case, while we opted for an online design, we wanted to ensure that the effects of tDCS on our participants' neurons would have started when they were performing the search task hence the use of the reaction time task as our filler activity. We also endeavoured to potentially avoid a fading effect of tDCS by opting for an online stimulation, as the effects on the neurons can be short-lived. Yet, having a filler task can cause an issue as highlighted by Horvath et al. [288] and Nozari et al. [292], they show that in many cases it can lead to the interference or to a vanishing effect of tDCS stimulation: Nozari et al. [292] designed a study where they used cathodal (inhibitory) stimulation on the left PFC at a current intensity of 1.5 mA. They manipulated the filler activity during stimulation (one related to the left PFC and with high cognitive demand, another which was a simple categorisation task with low cognitive demand): They found that when participants performed a filler task with low cognitive demand their participants' performance in the Flanker task (post-stimulation) improved while the performance post the cognitively demanding task decreased. They argue that tDCS effect could be moderated by the state of neuronal activation when participants start performing the actual task. In our case our filler task was a reaction time task which should involve the left DLPFC. Therefore the left DLPFC is mobilised and we should have found a tDCS effect on our exploratory search task. However, the nature of our experiment was clear to our participants (focused on search), hence perhaps that a potential lack of engagement during the reaction time task led to poor neuronal activation when we shifted to the exploratory search task. In addition, work by Quartarone et al. [293] has shown that anodal tDCS stimulation effect in an offline design vanished in the post stimulation task when participants were performing a task related to the post stimulation task, but in the cathodal tDCS stimulation condition, tDCS stimulation effect persisted.

Our finding related to participants correctly guessing their condition throws doubt on the efficacy of the sham condition as a way to achieve tDCS blinding, and this complements findings from previous work by Learmonth et al. [294]. The inability to achieve tDCS blinding can result in unwarranted consequences such as confounding factors where participants will adjust their behaviour based on the expectation of the stimulation [287].

Regarding the duration parameter, some studies have had smaller duration of tDCS stimulation and have been able to achieve positive results, therefore, longer duration does not necessarily lead to significant effects. It could also be that we did not stimulate for long enough and that lead to poor hyper-polarisation of the neurons in the left DLPFC. Similarly, to the duration parameter, a current intensity increase does not necessarily lead to higher efficacy, a current of 4 mA can fail to lead to an improvement [295]. Comparing an anodal tDCS stimulation of 1 mA against an anodal tDCS stimulation of 2 mA, a study has found significant working memory improvement only under 1 mA [296]. This variation of parameters had already been extensively discussed in past studies with tDCS, it adds to the need for future successful tDCS studies to be replicated, which might eventually lead on an agreement on the right parameters. This works adds to a growing literature showing the lack of effects of tDCS [266, 294, 297].

Our study stimulated the left DLPFC based on the use of an EEG cap, the generic dimension of such approach is sometimes considered controversial due to the variations in head dimensions which would lead to slight differences in the positioning of target brain regions, in this case the left DLPFC. However, our electrodes were large enough that the changes in position of a brain region by millimeters should not have significantly impacted the tDCS effect. In addition, past work in cognitive neuroscience has shown additional brain regions involved in executive functions and working memory, in parallel to the work in NeuraSearch which showed various brain regions were involved in the search process [124]: Therefore, it could be that the stimulation of only the left DLPFC was not enough to lead to improvement across our search task, it could also be that it was not the most effective area to target.

In addition to the tDCS-based parameter issues, a potential factor which could have had an impact on our results is the lack of screening based on the individuals' cognitive abilities. Previous work looking at individual factors has found that working memory capacity can influence language comprehension [270], work has also been done showing individual difference in selective attention based on the individuals' lifestyle [277]. A screening based on the cognitive control abilities could have led in having two groups (high WM span/low WM span) to see if there is an effect of working memory capacity when tDCS stimulation is performed on an exploratory search task.

Another potential reason behind our null results could be related to the topics offered. We selected 7 topics in order to offer a diverse range; however, we ended up with an irregularly distributed number of topics done across conditions. For example, topic 6, which was about Sudan after the December 2018 protests, was not done across the four conditions (Figure 4.11b).

4.4.2 Summary

In this chapter, we used a non-invasive brain stimulation method namely tDCS to stimulate the left DLPFC, a region associated with working memory. Our pre-registration hypothesised a reduction of the dwell time as a result of anodal tDCS stimulation, however our hypothesis was not supported. Additional outcome measures were identified and we did not find a tDCS effects for these either in an exploratory analysis. An analysis using linear mixed models highlighted previously noticed patterns of tDCS such as inter-subject variability. In addition we found that the feeling of side effects, for some of the outcome measures, had a moderating effect on participants' behaviour in some cases with and without an interaction with tDCS stimulation. In our discussion, we highlighted various factors that could have impacted our results and use our findings complemented with the literature to highlight the potential limits of tDCS and areas that should be paid careful attention when using tDCS.

Chapter 5

Conclusion

In this chapter, we start by summarising the main contributions of this thesis. Based on the findings of this thesis, we offer directions for future work and associated background, we set new research questions and potential steps to be taken to answer them: These research questions aim to contribute to coming closer to the complete disambiguation of information overload (IO) and to solving IO. Finally, as the concept of IO has recently taken an important place in individuals' lives through the excessive use of technology, and as it impacts their attention, we share recommendations for the research community, engineers, and policy-makers on how to approach it based on our findings.

5.1 Contributions

Information overload has long been present throughout history, however, we have seen that its forms have changed and, with the boom that the IR community and global computer and information science (CIS) community have seen in the development of new information systems, new dimensions of IO have appeared. This thesis was born from the observation of the ever-growing development of advanced recommender systems which were meant to support users and tackle IO by offering advanced filtering mechanisms. While IO was constantly mentioned, it always appeared in the CIS literature without a clear definition of which of IO's dimensions are tackled: It is treated as a simple problem but as demonstrated in chapter 2, it comes in various forms and affects various aspects of the individuals.

The first contribution of this thesis is the definition of IO and a clear framework presenting its various causes, manifestations, and consequences.

Based on the definition and framework of chapter 2 [9], we evaluated our framework by using the causes of IO we had found. In chapter 3, we were able to create IO in a controlled environment with a realistic task. The choice of a social network site (SNS), and in this case Twitter, was motivated by the relevance SNS have taken into our lives. Twitter has become a source of information for everyone, from your neighbour to your local authorities. Twitter, like other SNS, has its advantages in contributing to the financial improvement of individuals and to political change [129]. However, it can become an addiction, an addiction where the users' attention is heavily impacted as well as the way they process information. Therefore, studies like ours around SNS exemplify the problem of IO at a good level. In our study, we found that when impacted with IO, our participants changed their behaviour, their interaction with the feed changed. We were able to derive from our log data an accurate estimation of these changes through explicit and implicit interactions with the system: We found that users like less and revisit older tweets less, but most important of all the mouse tracking data is a key predictor to evaluate when users are under IO. We show the mouse tracking can be used as a reliable and scalable method to detect when individuals are under IO. This finding is relevant to the development of intelligent adaptive systems which can support individuals when they are subject to IO.

Finally, as we highlighted the current limits of system-side improvements to tackle IO. In chapter 4, we explored the problem of IO from a cognitive neuroscience perspective, and contribute to a growing body of work which uses neurophysiological tools in the field of IR. Our most ambitious work consisted of using neuromodulation to improve individuals' cognitive abilities. We used transcranial direct current stimulation (tDCS), a non-invasive brain stimulation technique, to target the left dorsolateral prefrontal cortex (DLPFC). A review of the previous work done in Neuro-Information Science and NeuraSearch showed the involvement of working memory during the search process [103, 124], a further review of the cognitive neuroscience literature allowed us to determine the important role of the prefrontal cortex in executive functions [243, 250], and in particular the important involvement of the left DLPFC in working-memory, attention, and decision-making. Our study consisted of search tasks on a clone of Google search engine. We hypothesised that an improvement to the left DLPFC would lead to a reduction of the dwell time (i.e time spent on reading web pages excluding the search engine results page): We believed that as participants would be able to retain more information in their working-memory, this would lead them to filter out faster irrelevant information. In addition, we derived additional outcome measures such as the SERP dwell time, and outcome measures related to the queries submitted as well as the web pages accessed. When comparing the active tDCS stimulation with the sham tDCS stimulation condition, we did not find any conclusive improvement when we looked at the various outcome measures. However, we found that there was inter-subject variability and that participants changed their behaviour based on the feeling of common tDCS side-effects (e.g tingling, headaches, itching etc). Both of theses findings have been previously discussed in the cognitive neuroscience literature around tDCS. To ensure tDCS is having an effect and is increasing neuronal activity: We can use EEG in conjunction with tDCS as it would allow to measure changes in amplitude and verify that tDCS did successfully stimulate the target region.

In addition, the reliability of tDCS has been questioned, we proposed for future research to consider other alternatives such as transcranial magnetic stimulation. Finally, an additional potential reason for the lack of tDCS effect could be that we targeted the wrong brain region, as highlighted in the literature other areas of the brain are involved with executive functions [244], and in the online search process [124]. We presented additional potential limitations in the discussion section of chapter 4.

Borrowing from cognitive load theory, we found that our work fits well within cognitive load theory and the types of cognitive loads. Sweller et al. [298] termed the three types as follows: intrinsic, extraneous, and germane. Intrinsic represents the load associated with the information itself, extraneous represents the way information is presented, and finally the germane consists of the representation of the information in the brain [298]. In our work, we focused on the way the information is presented in chapter 3, then we looked at supporting the cognitive process of the users of a search system by stimulating their working memory chapter 4.

In the next section, we present short term and long term potential work which can further contribute to the disambiguation of the concept of IO. A necessary disambiguation which would then lead to solutions to tackle IO.

5.2 Future work

In this section, we distinguish between short-term research objectives and long-term ones which are more ambitious.

Short term

A first set of studies should focus on how the additional triggers of IO [9], other than the quantity and speed of information, can be manipulated. In particular quality and diversity of information, to understand whether these characteristics of information lead to the same behavioural changes. A first step would be to create datasets similar to those used within the TREC community [1]. One dataset can be a series of topics with associated web search results which have been rated for the quality of the information presented and diversity among the rest of search results. Another dataset could be based on tweets for studies based on SNS, this dataset should allow for studies to be designed in the context of SNS: The tweets could have an associated rating of their relevance to the rest of the tweets, a diversity rating, and quality rating. These datasets can be used to design tasks in controlled environments to identify and measure the behavioural changes of individuals.

As we explore more triggers identified in chapter 2 [9], we should look whether all the triggers regardless of their type (information characteristics, working environment etc) lead to the same behavioural changes. As we identify ways to create IO in the lab, a close attention should be paid to the behavioural changes that result from IO. In order to achieve this, studies can use our findings from the outcome measures of our Twitter-

[1]The Text REtrieval Conference makes available various datasets for different types of tasks: https://trec.nist.gov/data.html

based study to identify whether the behavioural changes are consistent regardless of the triggers, e.g Does the mean fixation time always increase when individuals are under IO?

For our classification model, we performed the IO and not IO classification across data captured at the end of the session: Future work should investigate the outcome measures that can be leveraged to enable an online classification model. A first set of work should look into developing efficient classifiers that can perform classification of the state of individuals during a session. Such classifier would fall under what is known as online machine learning [299]. Our current work has shown that we could perform classification over the whole session to detect IO, however, looking at online machine learning would allow the development of rapid adaptive systems.

At an individual level, we should look into investigating the brain activations involved during IO. This research question can be approached from different contexts, but based on our work two relevant context are the context of search engines and one is the context of SNS. A future study could reproduce our Twitter-based study and use EEG with source localisation, which we highlighted as a cheaper mechanism than fMRI that can achieve high temporal and spatial resolution. Such future study should also aim to identify how the phases of IO highlighted in chapter 3 translate in terms of neural activity.

As we look at the individual level, we should also consider the knowledge gain dimension of IO, i.e What is the impact of IO on knowledge gain? A potential study should use objective and subjective measures to assess knowledge gain of individuals' [147] when they are under IO relative to when they are not. A subsequent study should look at individual differences in knowledge gain when IO is experienced.

As the cognitive neuroscience literature has highlighted, there are individual difference in the way individuals can direct their spatial attention [277], in addition to the working memory capacity [270]. Hence, we believe focus should be made into identifying which individual differences could lead to overcoming IO. Based on the previous research question, a future study should look at individuals differences related to the executive functions: Could individual factors such as lifestyle influence knowledge gain when individuals are under IO? Previous work has shown that heavy multi-taskers performed worse in task-switching [300] In addition, such individual differences should be taken into account when working with neuromodulation as it could moderate the effect of devices such as tDCS.

Part of our semi-structured interviews has shown strategy adoption in chapter 3. Our findings have also shown the adoption of various strategies by our participants in chapter 3 when under IO. Future work should look into investigating how these strategies compare and whether some strategies are better than others. Another subsequent question is: “What are the elements that contribute to the development of these strategies?”, it should be investigated whether these strategies are a result of practice or unconscious adaptive behaviour that could be moderated by the ability of individuals to direct their attention. Another question can be derived based on the work by Savolainen [74] regarding strategies: ‘What is the predominant type of strategy adopted when under IO?’. Such questions should look into highlighting whether individuals are more prone to emergent or deliberate strategies, future work should also look at whether IO could hinder the development of either of those types of strategies.

One limitation we have highlighted as part of our Twitter-based study (chapter 3) is that we focused on text, since we did not display tweets with images and did not allow our participants to explore external websites. A study should look into the different modalities available (images, videos, sound) and whether we can quantify how much of each of modality leads to IO, it should question whether some modalities can carry more information and still not lead to IO. This is particularly relevant as we shift towards the development of voice assistants, some work has already started to look into the cognitive load generated by integrating with voice assistants [301].

As tablets become a predominant alternative for personal computers, in the future, we should explore efficient implicit and explicit feedback systems which can be used on smartphones and tablets to detect IO. A study should design new clones of real-world systems where IO could be created, this will allow to study potential implicit and explicit feedback systems to detect individuals’ behavioural changes on different mediums.

Long term

As we progress towards the disambiguation of IO, in the long term we should develop automated evaluation techniques of the individuals performance in user studies. For example, these evaluation methods should measure the individuals' knowledge change. This could be achieved as we continue progressing in the developments of language models, these were shown to be able to act as knowledge base [302]. Hence future work could look at evaluation methods of knowledge change based on language models, as this would allow faster progress in research and would allow to not rely on the subjectivity of annotators. Another long term goal should explore how we can create a working environment [9] in which we can control IO. One approach could be to leverage the progress made in virtual reality. Work in the field of virtual reality is constantly being done to increase the sense of immersion of the users [303].

Finally, future work should adopt a proactive approach and seek to identify how new technologies could bring new causes of IO. By having a sense of foresight, we can effectively avoid the further development of IO.

5.3 Broader impact and recommendations

Our concept analysis comes at an important time as policy-makers become more aware of the impact of the use of technology in society [15], as well as companies becoming aware that individuals are becoming more empowered and will leave platforms that lead them to a state poor well-being [145]. It offers a clearly defined framework with the dimensions of IO which can allow a better auditing of technology companies, in particular social network sites. Through a Twitter study, we were able to show the importance of such IO framework: We were able to create IO in a controlled environment and extract from behavioural data what can be used to detect the state of IO. This can be leveraged by engineers when developing these systems. The development of systems which can integrate these individuals insights on the users' state can be beneficial particularly for populations that are not used to using technology: Having intelligent adaptive systems which can offer support to less experienced users can lead to a more continuous use but also to better well-being. While technology companies have to ensure that users' stay on their platform, as this allows them to have users facing more monetised content, we believe we can find a good balance where platforms can gain more users while offering a user-system healthy relationship.

In addition, in the field of IR and interactive information retrieval (IIR), we offer log data which can be used to further model individuals' behaviour and integrate the IO dimension to their models. As we also contributed the code to our systems, we think this can lead to future studies which can use the same outcome measures hence reinforcing our findings and offering valuable insights to the industry. Being able to use similar outcome measures and derive them from the same environment (i.e system, screen resolution etc.) will lead to more robust contributions.

For researchers who are looking into exploring further the field of IR/IIR and CIS in general from a neuroscience perspective, we recommend a closer collaboration with researchers in cognitive neuroscience when designing their user studies. It can be tempting to want to perform ambitious user studies where various cognitive mechanisms might be involved, however, such approach might lead to accidental findings which occur only due to various confounding factors. Particularly in the case of brain stimulation, user studies should be restricted to one outcome measure. As we have seen in our exploratory analysis, we did find changes in behaviour in specific outcome measures when participants felt a tDCS side effect and a modulating effect of tDCS stimulation, however such findings would require further work focused on those specific outcome measures to confirm them.

Finally, as we highlighted in the section 4.4, it is important to keep in mind the ethical implications of these devices. If we are able to keep moving forward and successfully integrate brain stimulation devices for tasks such as search, or interacting with social networks sites to filter information, we need to understand the implications of such work. Hence future work should discuss how these techniques can be used, who it could profit, and how it can be shared with the general population.

Bibliography

[1] G. Marchionini and R. White, “Find what you need, understand what you find,” *International Journal of Human-Computer Interaction*, vol. 23, no. 3, pp. 205–237, 2007.

[2] B. L. Rodgers, “Concepts, analysis and the development of nursing knowledge: The evolutionary cycle,” *Journal of Advanced Nursing*, vol. 14, no. 4, pp. 330–335, 1989.

[3] D. S. Robertson, “The information revolution,” *Communication Research*, vol. 17, no. 2, pp. 235–254, 1990.

[4] L. Floridi, *The fourth revolution: How the infosphere is reshaping human reality.* OUP Oxford, 2014.

[5] M. Sanderson and W. B. Croft, “The History of Information Retrieval Research,” *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1444–1451, 2012.

[6] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer Networks and ISDN Systems*, vol. 30, no. 1, pp. 107–117, 1998, proceedings of the Seventh International World Wide Web Conference. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016975529800110X

[7] S.-O. Lee, A. Hong, and J. Hwang, “ICT diffusion as a determinant of human progress,” *Information Technology for Development*, vol. 23, no. 4, pp. 687–705, 2017.

[8] D. Bawden and L. Robinson, “The dark side of information: Overload, anxiety and other paradoxes and pathologies,” *Journal of Information Science*, vol. 35, no. 2, pp. 180–191, 2009.

[9] M. A. Belabbes, I. Ruthven, Y. Moshfeghi, and D. R. Pennington, “Information overload: A concept analysis,” *Journal of Documentation*, no. ahead-of-print, 2022.

[10] L. Wu, Z. Zhu, H. Cao, and B. Li, “Influence of information overload on operator’s user experience of human–machine interface in led manufacturing systems,” *Cognition, Technology & Work*, vol. 18, no. 1, pp. 161–173, 2015.

[11] N. J. Belkin, “Information concepts for information science,” *Journal of Documentation*, 1978.

[12] P. Vakkari, “Information seeking in context: A challenging metatheory,” in *Proceedings of an international conference on Information Seeking in Context*, 1997, pp. 451–464.

[13] K. Jarvelin and P. Vakkari, “Content analysis of research articles in library and information science.” *Library & Information Science Research*, vol. 12, no. 4, pp. 395–421, 1990.

[14] R. A. Fleming-May, “Concept analysis for library and information science: Exploring usage,” *Library & Information Science Research*, vol. 36, no. 3-4, pp. 203–210, 2014.

[15] S. Donaldson and C. Mai-Duc, “Meta, TikTok Could Face Civil Liability if Held to Addict Children in California,” The Wall Street Journal, June 19 2022. [Online]. Available: https://www.wsj.com/articles/meta-tiktok-could-face-civil-liability-for-addicting-children-in-california-11656419401

[16] A. S. Foley and A. H. Davis, “A guide to concept analysis,” *Clinical Nurse Specialist*, vol. 31, no. 2, pp. 70–73, 2017.

[17] V. Navalpakkam and E. Churchill, “Mouse tracking: Measuring and Predicting Users’ Experience of Web-Based Content,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’12. New York, NY, USA: Association for Computing Machinery, 2012, p. 2963–2972. [Online]. Available: https://doi.org/10.1145/2207676.2208705

[18] R. W. White and R. A. Roth, “Exploratory search: Beyond the query-response paradigm,” *Synthesis lectures on information concepts, retrieval, and services*, vol. 1, no. 1, pp. 1–98, 2009.

[19] V. Bush, “As we may think,” *The Atlantic Monthly*, vol. 176, no. 1, pp. 101–108, 1945.

[20] D. Bawden, C. Holtham, and N. Courtney, “Perspectives on information overload,” *Aslib Proceedings*, vol. 51, no. 8, pp. 249–255, 1999.

[21] J. Sweller, “Cognitive load during problem solving: Effects on learning,” *Cognitive Science*, vol. 12, no. 2, pp. 257–285, 1988.

[22] J. MacDonald, P. Bath, and A. Booth, “Information overload and information poverty: Challenges for healthcare services managers?” *Journal of Documentation*, vol. 67, no. 2, pp. 238–263, 2011.

[23] K. Koroleva, H. Krasnova, and O. Günther, “Stop spamming me!: Exploring information overload on Facebook,” in *Proceedings of the Americas Conference on Information Systems (AMCIS)*, 2010.

[24] Y. Sasaki, D. Kawai, and S. Kitamura, “The anatomy of tweet overload: How number of tweets received, number of friends, and egocentric network density affect perceived information overload,” *Telematics and Informatics*, vol. 32, no. 4, pp. 853–861, 2015.

[25] S. Fu, H. Li, Y. Liu, H. Pirkkalainen, and M. Salo, “Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload,

system feature overload, and social overload," *Information Processing & Management*, vol. 57, no. 6, pp. 102–307, 2020.

[26] T. Koltay, "Information overload in a data-intensive world," in *Understanding Information*. Springer, 2017, pp. 197–217.

[27] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, 1972.

[28] M. Montebello, "Information overload: An IR problem?" in *Proceedings of String Processing and Information Retrieval: A South American Symposium*. IEEE, 1998, pp. 65–74.

[29] S. R. Hiltz and L. Plotnick, "Dealing with information overload when using social media for emergency management: Emerging solutions," in *Proceedings of the 10th International ISCRAM Conference*, 2013, pp. 823–827.

[30] L. Lauri and S. Virkus, "Information Overload of Academic Staff in Higher Education Institutions in Estonia," in *European Conference on Information Literacy*. Springer, 2018, pp. 347–356.

[31] P. G. Rötzel and D. D. Fehrenbacher, "On the role of information overload in information systems (IS) success: Empirical evidence from decision support systems," in *Proceedings of the 2019 International Conference on Information Systems (ICIS)*, 2019.

[32] D. Bawden and L. Robinson, "Information overload: An introduction," in *Oxford encyclopedia of political decision making*. Oxford University Press, 2021.

[33] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, 2014, pp. 739–748.

[34] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, "A review on deep learning

for recommender systems: Challenges and remedies," *The Artificial Intelligence Review*, vol. 52, no. 1, pp. 1–37, 2019.

[35] M.-A. Kaufhold, N. Rupp, C. Reuter, and M. Habdank, "Mitigating information overload in social media during conflicts and crises: Design and evaluation of a cross-platform alerting system," *Behaviour & Information Technology*, vol. 39, no. 3, pp. 319–342, 2020.

[36] C. Zerva, M.-Q. Nghiem, N. T. Nguyen, and S. Ananiadou, "Cited text span identification for scientific summarisation using pre-trained encoders," *Scientometrics*, vol. 125, no. 3, pp. 3109–3137, 2020.

[37] H. Costa and L. Macedo, "Emotion-based recommender system for overcoming the problem of information overload," *Highlights on Practical Applications of Agents and Multi-Agent Systems*, vol. 365, pp. 178–189, 2013.

[38] Y. Moshfeghi and J. M. Jose, "An effective implicit relevance feedback technique using affective, physiological and behavioural features," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 2013, pp. 133–142.

[39] D. R. Pennington, "“The most passionate cover I've seen”: Emotional information in fan-created U2 music videos," *Journal of Documentation*, vol. 72, no. 3, pp. 569–590, 2016.

[40] M. J. Eppler and J. Mengis, "The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines," in *Kommunikationsmanagement im Wandel.* Springer, 2008, pp. 271–305.

[41] C. Gunaratne, N. Baral, W. Rand, I. Garibay, C. Jayalath, and C. Senevirathna, "The effects of information overload on online conversation dynamics," *Computational and Mathematical Organization Theory*, vol. 26, no. 2, pp. 255–276, 2020.

[42] J. Matthes, K. Karsay, D. Schmuck, and A. Stevic, "Too much to handle: Impact of mobile social networking sites on information overload, depressive symptoms, and well-being," *Computers in Human Behavior*, vol. 105, pp. 106–217, 2020.

[43] G. Phillips-Wren and M. Adya, "Decision making under stress: The role of information overload, time pressure, complexity, and uncertainty," *Journal of Decision Systems*, vol. 29, no. 1, pp. 213–225, 2020.

[44] R. Tofthagen and L. M. Fagerstrøm, "Rodgers' evolutionary concept analysis–a valid method for developing knowledge in nursing science," *Scandinavian Journal of Caring Sciences*, vol. 24 (s1), pp. 21–31, 2010.

[45] E. Whelan and R. Teigland, "Transactive memory systems as a collective filter for mitigating information overload in digitally enabled organizational groups," *Information and Organization*, vol. 23, no. 3, pp. 177–197, 2013.

[46] B. Jacobfeuerborn and M. Muraszkiewicz, "Media, information overload, and information science," *Intelligent Tools for Building a Scientific Information Platform*, vol. 467, pp. 3–13, 2013.

[47] D. Memmi, "Information overload and virtual institutions," *AI & Society*, vol. 29, no. 1, pp. 75–83, 2014.

[48] A. Ndumu, "Toward a new understanding of immigrant information behavior: A survey study on information access and information overload among US Black diasporic immigrants," *Journal of Documentation*, vol. 76, no. 4, pp. 869–891, 2020.

[49] D. Diderot *et al.*, *Encyclopédie Ou Dictionnaire Raisonné Des Sciences, Des Arts Et Des Métiers: Do-Esy. 5.* Paris, France: Briasson, 1755, vol. 5.

[50] G. M. Beard, *American nervousness, its causes and consequences: A supplement to nervous exhaustion (neurasthenia).* G.P. Putnam's Sons, 1881.

[51] D. G. Schuster, "Neurasthenia and a Modernizing America," *JAMA*,

vol. 290, no. 17, pp. 2327–2328, 11 2003. [Online]. Available: https://doi.org/10.1001/jama.290.17.2327

[52] D. M. Levy, "Information overload," in *The Handbook of Information and Computer Ethics*, 2008, pp. 497–515.

[53] R. L. Meier, *A communications theory of urban growth.* Joint Center for Urban Studies of the Massachusetts Institute of Technology, 1962.

[54] B. M. Gross, "The managing of organizations: The administrative struggle," *Administrative Law Review*, vol. 19, no. 1, pp. 197–198, 1965.

[55] A. Toffler, *Future Shock.* New York: Random House, 1970.

[56] S. Virkus, S. Mandre, and E. Pals, "Information overload in a disciplinary context," in *Proceedings of the European Conference on Information Literacy (ECIL) 2017*, 2017, pp. 615–624.

[57] G. A. Miller, "The magic number seven plus or minus two: Some limits on our capacity for processing information," *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.

[58] H. M. Schroder, S. Streufert, and M. J. S. Driver, *Human Information Processing: Individuals and Groups Functioning in Complex Social Situations.* New York: Holt, Rinehart and Winston, 1967.

[59] R. Savolainen, "Filtering and withdrawing: Strategies for coping with information overload in everyday contexts," *Journal of Information Science*, vol. 33, no. 5, pp. 611–621, 2007.

[60] C.-Y. Chen, S. Pedersen, and K. L. Murphy, "Learners' perceived information overload in online learning via computer-mediated communication," *Research in Learning Technology*, vol. 19, no. 2, pp. 101–116, 2011.

[61] H. Shrivastav and S. Hiltz, "Information overload in technology-based education: A meta-analysis," *19th Americas Conference on Information Systems, AMCIS*

2013 - Hyperconnected World: Anything, Anywhere, Anytime, vol. 3, pp. 1994–2003, 01 2013.

[62] M. G. Rodriguez, K. Gummadi, and B. Schoelkopf, "Quantifying information overload in social media and its impact on social contagions," in *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, vol. 8, no. 1, 2014.

[63] E. M. Hallowell, "Overloaded circuits: Why smart people underperform," *Harvard business review*, vol. 83, no. 1, pp. 54–62, 2005.

[64] D. Kirsh, "A few thoughts on cognitive overload," *Intellectica*, vol. 1, no. 30, 2000.

[65] H. Liang and K.-W. Fu, "Information overload, similarity, and redundancy: Unsubscribing information sources on Twitter," *Journal of Computer-Mediated Communication*, vol. 22, no. 1, pp. 1–17, 2017.

[66] A. Keselman, G. Rosemblat, H. Kilicoglu, M. Fiszman, H. Jin, D. Shin, and T. C. Rindflesch, "Adapting semantic natural language processing technology to address information overload in influenza epidemic management," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2531–2543, 2010.

[67] S. Zhang, "Information overload and online collaborative learning: Insights from agent-based modeling," in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, 2018, pp. 1–2.

[68] M. d. C. Ocón Palma, A.-M. Seeger, and A. Heinzl, "Mitigating information overload in e-commerce interactions with conversational agents," in *Information Systems and Neuroscience*, F. D. Davis, R. Riedl, J. vom Brocke, P.-M. Léger, A. Randolph, and T. Fischer, Eds. Cham: Springer International Publishing, 2020, pp. 221–228.

[69] A. E. Holton and H. I. Chyi, "News and the overloaded consumer: Factors influencing information overload among news consumers," *Cyberpsychology, Behavior, and Social Networking*, vol. 15, no. 11, pp. 619–624, 2012.

[70] J. B. Schmitt, C. A. Debbelt, and F. M. Schneider, "Too much information? Predictors of information overload in the context of online news exposure," *Information, Communication & Society*, vol. 21, no. 8, pp. 1151–1167, 2018.

[71] Z. A. Sabeeh and Z. Ismail, "Effects of information overload on productivity in enterprises: A literature review," in *Proceedings of the 2013 International Conference on Research and Innovation in Information Systems (ICRIIS)*. IEEE, 2013, pp. 210–214.

[72] R. P. F. Marques and J. C. L. Batista, "An overview on information and communication overload," in *Information and communication overload in the digital age*. IGI Global, 2017, pp. 1–19.

[73] T. W. Jackson and P. Farzaneh, "Theory-based model of factors affecting information overload," *International Journal of Information Management*, vol. 32, no. 6, pp. 523–532, 2012.

[74] R. Savolainen, "Information seeking and searching strategies as plans and patterns of action: A conceptual analysis," *Journal of Documentation*, 2016.

[75] ——, "Conceptualizing information need in context," *Information Research*, vol. 17, 12 2012.

[76] S. D. Lambert and C. G. Loiselle, "Health information—seeking behavior," *Qualitative health research*, vol. 17, no. 8, pp. 1006–1019, 2007.

[77] M. S. Zimmerman and G. Shaw Jr, "Health information seeking behaviour: A concept analysis," *Health Information & Libraries Journal*, vol. 37, no. 3, pp. 173–191, 2020.

[78] R. A. Fleming-May, "What is library use? Facets of concept and a typology of

its application in the literature of library and information science," *The Library Quarterly*, vol. 81, no. 3, pp. 297–320, 2011.

[79] R. Savolainen, "Cognitive barriers to information seeking: A conceptual analysis," *Journal of Information Science*, vol. 41, no. 5, pp. 613–623, 2015.

[80] S. Maidullah and A. Sharma, "Decision making in the era of infobesity: A study on interaction of gender and psychological tendencies," *Humanities & Social Sciences Reviews*, vol. 7, no. 5, pp. 571–586, 2019.

[81] C. Luo, Y. Lan, C. Wang, and L. Ma, "The effect of information consistency and information aggregation on eWOM readers' perception of information overload," in *Proceedings of the Pacific Asia Conference on Information Systems (PACIS) 2013*, 2013, p. 180.

[82] M. Tzagarakis, S. Christodoulou, and N. Karacapilidis, "Alleviating cognitive overload in collaboration support systems," *Frontiers in Artificial Intelligence and Applications*, vol. 262, pp. 172–181, 2014.

[83] G. Poghosyan, "Addressing information overload through text mining across news and social media streams," in *Proceedings of the 5th International Workshop on Social Media World Sensors*, 2019, pp. 1–2.

[84] S. Dutta, "Actionable filtering themes in executive information systems for managing information overload," *Journal of Theoretical & Applied Information Technology*, vol. 26, pp. 16–21, 2011.

[85] A. Lincoln, "FYI: TMI: Toward a holistic social theory of information overload," *First Monday*, vol. 16, no. 3, 2011.

[86] G. Y.-M. Kao and C.-C. Peng, "A multi-source book review system for reducing information overload and accommodating individual styles," *Library Hi Tech*, vol. 33, no. 3, pp. 310–328, 2015.

[87] J. C. Benselin and G. Ragsdell, "Information overload: The differences that age

makes," *Journal of Librarianship and Information Science*, vol. 48, no. 3, pp. 284–297, 2016.

[88] P. Katz, T. Lunze, M. Feldmann, D. Röhrborn, and A. Schill, "System architecture for handling the information overload in enterprise information aggregation systems," *Business Information Systems*, vol. 87, pp. 148–159, 2011.

[89] G. Marchionini, "Exploratory search: From finding to understanding," *Communications of the ACM*, vol. 49, no. 4, pp. 41–46, 2006.

[90] B. Swar, T. Hameed, and I. Reychav, "Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search," *Computers in Human Behavior*, vol. 70, pp. 416–425, 2017.

[91] E. Simperl, I. Thurlow, P. Warren, F. Dengler, J. Davies, M. Grobelnik, D. Mladenić, J. M. Gómez-Pérez, and C. R. Moreno, "Overcoming information overload in the enterprise: The active approach," *IEEE Internet Computing*, vol. 14, no. 6, pp. 39–46, 2010.

[92] A.-F. Rutkowski and C. S. Saunders, "Growing pains with information overload," *Computer*, vol. 43, pp. 94–96, 2010.

[93] J. B. Strother, J. M. Ulijn, and Z. Fazal, "Effect of Color, Visual Form, and Textual Information on Information Overload," in *Information Overload: An International Challenge for Professional Engineers and Technical Communicators*. Wiley-IEEE Press, 2012, pp. 103–124.

[94] F. A. D'Asaro, V. Perticone, and M. E. Tabacchi, "A fuzzy methodology to alleviate information overload in eLearning," in *Proceedings of the 8th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-13)*. Atlantis Press, 2013.

[95] E. Whelan and R. Teigland, "Managing Information Overload: Examining the Role of the Human Filter," in *Proceedings of the 9th European Conference on Information Systems (ECIS) 2011*, 2011.

[96] S. Yang and A. Albers, "Overcoming information overload in online reputation management: A systematic literature review," in *Proceedings of the European Conference on Information Systems (ECIS) 2013*, 2013.

[97] R. K. Minas and M. E. Crosby, "Overloaded and biased? Using augmented cognition to understand the interaction between information overload and cognitive biases," in *Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience (AC 2016). Lecture Notes in Computer Science*, 2016, vol. 9743, pp. 242–252.

[98] D. Saxena and M. Lamest, "Information overload and coping strategies in the big data context: Evidence from the hospitality sector," *Journal of Information Science*, vol. 44, no. 3, pp. 287–297, 2018.

[99] E. Hargittai, W. R. Neuman, and O. Curry, "Taming the information tide: Perceptions of information overload in the American home," *The Information Society*, vol. 28, no. 3, pp. 161–173, 2012.

[100] N. Abdullah and N. A. Mustapar, "Heart rate reading patterns of moment of information overload during online information searching," *Journal of Information and Communication Technology*, vol. 18, no. 3, pp. 233–251, 2019.

[101] Y. Ouardi, T. Goyal, L. Graf-Vlachy, J. Mammen, A. König, and C. Saunders, "The cost of sharing: The effect of sharing inclination on information overload," in *European Conference on Information Systems. Research Papers*, vol. 162, 2016.

[102] S. Paul and D. L. Nazareth, "Input information complexity, perceived time pressure, and information processing in GSS-based work groups: An experimental investigation using a decision schema to alleviate information overload conditions," *Decision Support Systems*, vol. 49, no. 1, pp. 31–40, 2010.

[103] P. Bera, "Does cognitive overload matter in understanding BPMN models?" *Journal of Computer Information Systems*, vol. 52, no. 4, pp. 59–69, 2012.

[104] N. A. Mustapar, N. Abdullah, and N. L. M. Noor, "A review towards developing a moment of information overload model," in *2016 4th International Conference on User Science and Engineering (i-USEr)*. IEEE, 2016, pp. 222–226.

[105] P. Melinat, T. Kreuzkam, and D. Stamer, "Information overload: A systematic literature review," in *Proceedings of the International Conference on Business Informatics Research*, 2014, pp. 72–86.

[106] J. B. Strother, J. M. Ulijn, and Z. Fazal, "Of Time Magazine, 24/7 Media, and Data Deluge: The Evolution of Information Overload Theories and Concepts," in *Information Overload: An International Challenge for Professional Engineers and Technical Communicators*. Wiley-IEEE Press, 2012, pp. 13–40.

[107] J. Tran, Q. V. Nguyen, A. Hol, and S. Simoff, "Reduction of Cognitive Overload in Online Reviews Using Data Visualisation," in *Proceedings of the Australasian Computer Science Week Multiconference*, 2019, pp. 1–9.

[108] U. Wild, S. Laumer, and A. Kroenke, "The Bipartite Influence of Information Overload on User Resistance to Knowledge Management Systems," in *Proceedings of the Americas Conference on Information Systems (AMCIS) 2012*, 2012.

[109] S.-C. Hu and I.-C. Chen, "Alleviating information overload caused by volumes of numerical web data: The concept and development process," *IET Software*, vol. 5, no. 5, pp. 445–453, 2011.

[110] A. Green, "Information overload in healthcare management: How the READ Portal is helping healthcare managers," *Journal of the Canadian Health Libraries Association (JCHLA)*, vol. 32, no. 3, pp. 173–176, 2011.

[111] W.-K. Tan and P.-C. Kuo, "The consequences of online information overload confusion in tourism," *Information Research*, vol. 24, no. 2, 2019.

[112] J. B. Strother, J. M. Ulijn, and Z. Fazal, *Drowning in Data: A Review of Information Overload within Organizations and the Viability of Strategic Communication Principles*. Wiley-IEEE Press, 2012, pp. 231–250.

[113] C. P. Furner and R. A. Zinko, “The influence of information overload on the development of trust and purchase intention based on online product reviews in a mobile vs. web environment: An empirical investigation,” *Electronic Markets*, vol. 27, no. 3, pp. 211–224, 2017.

[114] Aviation Management, “Information overload: The next technological challenge,” in *Proceedings of the 2011 Integrated Communications, Navigation, and Surveillance Conference*, 2011.

[115] S. Talwar, A. Dhir, D. Singh, G. Virk, and J. Salo, “Sharing of fake news on social media: Application of the honeycomb framework and the third-person effect hypothesis,” *Journal of Retailing and Consumer Services*, vol. 57, p. 102197, 11 2020.

[116] S. Dennis, P. Bruza, and R. McArthur, “Web searching: A process-oriented experimental study of three interactive search paradigms,” *Journal of the American Society for Information Science and Technology*, vol. 53, no. 2, pp. 120–133, 2002.

[117] Y.-M. Kim and S. Y. Rieh, “Dual-task performance as a measure for mental effort in library searching and web searching,” in *Proceedings of the 68th annual meeting of the American society for information science and technology*, vol. 42, 2005.

[118] J. Gwizdka, “Cognitive Load and Web Search Tasks,” *HCIR*, 2009.

[119] S. Harper, E. Michailidou, and R. Stevens, “Toward a definition of visual complexity as an implicit measure of cognitive load,” *ACM Transactions on Applied Perception (TAP)*, vol. 6, no. 2, pp. 1–18, 2009.

[120] P. Schmutz, S. Heinz, Y. Métrailler, and K. Opwis, “Cognitive load in ecommerce applications—measurement and effects on user satisfaction,” *Advances in Human-Computer Interaction*, vol. 2009, 2009.

[121] S. Chen, J. Epps, N. Ruiz, and F. Chen, “Eye activity as a measure of human mental effort in HCI,” in *Proceedings of the 16th international conference on Intelligent user interfaces*, 2011, pp. 315–318.

[122] J. R. Stroop, "Studies of interference in serial verbal reactions." *Journal of experimental psychology*, vol. 18, no. 6, p. 643, 1935.

[123] F. Chen, N. Ruiz, E. Choi, J. Epps, M. A. Khawaja, R. Taib, B. Yin, and Y. Wang, "Multimodal behavior and interaction as indicators of cognitive load," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 4, pp. 1–36, 2013.

[124] Y. Moshfeghi and F. E. Pollick, "Search process as transitions between neural states," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 1683–1692.

[125] W. Jo, S. S. Kannan, G.-E. Cha, A. Lee, and B.-C. Min, "Rosbag-based multimodal affective dataset for emotional and cognitive states," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020, pp. 226–233.

[126] G.-E. Cha and B.-C. Min, "Correlation between Unconscious Mouse Actions and Human Cognitive Workload," in *CHI'22 Conference on Human Factors in Computing Systems Extended Abstracts*, 2022, pp. 1–7.

[127] M. R. Auer, "The policy sciences of social media," *Policy Studies Journal*, vol. 39, no. 4, pp. 709–736, 2011.

[128] N. Erskine, S. Hendricks *et al.*, "The use of Twitter by medical journals: Systematic review of the literature," *Journal of medical Internet research*, vol. 23, no. 7, p. e26378, 2021.

[129] W. O. Anyim, "Twitter Ban in Nigeria: Implications on Economy, Freedom of Speech and Information Sharing," *Library Philosophy and Practice*, pp. 0_1–13, 2021.

[130] T. C. Helmus and E. Bodine-Baron, *Empowering ISIS opponents on Twitter*. Rand Corporation, 2017.

[131] M. C. Galgoczy, A. Phatak, D. Vinson, V. K. Mago, and P. J. Giabbanelli, "(Re) shaping online narratives: When bots promote the message of President Trump during his first impeachment," *PeerJ Computer Science*, vol. 8, p. e947, 2022.

[132] Cambridge Dictionary, "Silent majority," in *Cambridge Dictionary*, 2022. [Online]. Available: https://dictionary.cambridge.org/dictionary/english/silent-majority

[133] N. Lu, G. Wu, Z. Zhang, Y. Zheng, Y. Ren, and K.-K. R. Choo, "Cyberbullying detection in social media text based on character-level convolutional neural network with shortcuts," *Concurrency and Computation: Practice and Experience*, vol. 32, no. 23, p. e5627, 2020.

[134] A. Bermes, "Information overload and fake news sharing: A transactional stress perspective exploring the mitigating role of consumers' resilience during covid-19," *Journal of Retailing and Consumer Services*, vol. 61, 07 2021.

[135] S. Zhang, L. Zhao, Y. Lu, and J. Yang, "Do you get tired of socializing? An empirical explanation of discontinuous usage behaviour in social network services," *Information & Management*, vol. 53, no. 7, pp. 904–914, 2016.

[136] R. S. Wurman, *Information Anxiety.* New York: Doubleday, 1989.

[137] R. Ayyagari, V. Grover, and R. Purvis, "Technostress: Technological antecedents and implications," *MIS quarterly*, pp. 831–858, 2011.

[138] N. O. Hodas and K. Lerman, "How visibility and divided attention constrain social contagion," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing.* IEEE, 2012, pp. 249–257.

[139] N. Koumchatzky and A. Andryeyev, "Using Deep Learning at Scale in Twitter's Timelines," May 9 2017. [Online]. Available: https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines

[140] N. Rubio-Licht, "Twitter abandons unpopular feed change," Protocol, March 14 2022. [Online]. Available: https://www.protocol.com/bulletins/twitter-chronological-timeline

[141] N. Bartley, A. Abeliuk, E. Ferrara, and K. Lerman, "Auditing algorithmic bias on Twitter," in *13th ACM Web Science Conference 2021*, 2021, pp. 65–73.

[142] A.-A. Stoica, C. Riederer, and A. Chaintreau, "Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity," in *Proceedings of the 2018 World Wide Web Conference*, 2018, pp. 923–932.

[143] E. Bakshy, S. Messing, and L. A. Adamic, "Exposure to ideologically diverse news and opinion on Facebook," *Science*, vol. 348, no. 6239, pp. 1130–1132, 2015.

[144] S. Flaxman, S. Goel, and J. M. Rao, "Filter bubbles, echo chambers, and online news consumption," *Public opinion quarterly*, vol. 80, no. S1, pp. 298–320, 2016.

[145] A. Hern, "Tiktok addicts to benefit from in-app screen-time controls," The Guardian, June 9 2022. [Online]. Available: https://www.theguardian.com/technology/2022/jun/09/tiktok-addicts-to-benefit-from-in-app-screen-time-controls

[146] J. Gwizdka, Y. Moshfeghi, and M. L. Wilson, "Introduction to the special issue on neuro-information science," *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, 2019.

[147] N. Bhattacharya and J. Gwizdka, "Relating eye-tracking measures with changes in knowledge on search tasks," in *Proceedings of the 2018 ACM symposium on eye tracking research & applications*, 2018, pp. 1–5.

[148] W. A. Wan Adnan, W. N. H. Hassan, N. Abdullah, and J. Taslim, "Eye tracking analysis of user behavior in online social networks," in *International Conference on Online Communities and Social Computing*. Springer, 2013, pp. 113–119.

[149] N. Nourbakhsh, F. Chen, Y. Wang, and R. A. Calvo, "Detecting Users' Cognitive Load by Galvanic Skin Response with Affective Interference," *ACM*

Trans. Interact. Intell. Syst., vol. 7, no. 3, sep 2017. [Online]. Available: https://doi.org/10.1145/2960413

[150] J. Gwizdka and A. Dillon, "Eye-tracking as a method for enhancing research on information search," in *Understanding and Improving Information Search*. Springer, 2020, pp. 161–181.

[151] L. Tonin, E. Menegatti, and D. Coyle, "Advances in the integration of brain-machine interfaces and robotic devices," p. 653615, 2021.

[152] N. Jaques, J. Engel, D. Ha, F. Bertsch, R. Picard, and D. Eck, "Learning via social awareness: improving sketch representations with facial feedback," in *International Conference on Learning Representations*, 2018, workshop Track. [Online]. Available: https://arxiv.org/abs/1802.04877

[153] S. B. Sura, N. M. Nistah, S. Lee, and D. B. Alebaba, "Using Eye Tracking Approach in Analyzing Social Network Site Area of Interest for Consumers' Decision Making in Social Commerce," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, 2021.

[154] L. Shi, N. Bhattacharya, A. Das, M. Lease, and J. Gwizdka, "The effects of interactive ai design on user behavior: An eye-tracking study of fact-checking covid-19 claims," in *ACM SIGIR Conference on Human Information Interaction and Retrieval*, 2022, pp. 315–320.

[155] M. A. Just and P. A. Carpenter, "A theory of reading: From eye fixations to comprehension." *Psychological review*, vol. 87, no. 4, p. 329, 1980.

[156] J. M. Findlay and I. D. Gilchrist, "Visual selection, covert attention and eye movements," *Active vision: The psychology of looking and seeing*, pp. 35–54, 2003.

[157] J. Gwizdka, "Characterizing relevance with eye-tracking measures," in *Proceedings of the 5th information interaction in context symposium*, 2014, pp. 58–67.

[158] J. Simko, M. Hanakova, P. Racsko, M. Tomlein, R. Moro, and M. Bielikova, "Fake news reading on social media: An eye-tracking study," in *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, 2019, pp. 221–230.

[159] S. Chen and J. Epps, "Automatic classification of eye activity for cognitive load measurement with emotion interference," *Computer methods and programs in biomedicine*, vol. 110, no. 2, pp. 111–124, 2013.

[160] T. Kosch, M. Hassib, P. W. Woźniak, D. Buschek, and F. Alt, "Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ser. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1–13. [Online]. Available: https://doi.org/10.1145/3173574.3174010

[161] G. G. M. Dalveren and N. E. Cagiltay, "Using eye-movement events to determine the mental workload of surgical residents," *Journal of Eye Movement Research*, vol. 11, no. 4, 2018.

[162] J. Zagermann, U. Pfeil, and H. Reiterer, "Measuring cognitive load using eye tracking technology in visual computing," in *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization*, 2016, pp. 78–85.

[163] K. Rodden, X. Fu, A. Aula, and I. Spiro, "Eye-mouse coordination patterns on web search results pages," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 2008, pp. 2997–3002.

[164] Q. Guo and E. Agichtein, "Towards predicting web searcher gaze position from mouse movements," in *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 2010, pp. 3601–3606.

[165] J. Huang, R. White, and G. Buscher, "User see, user point: Gaze and cursor alignment in web search," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 1341–1350.

[166] V. Navalpakkam, L. Jentzsch, R. Sayres, S. Ravi, A. Ahmed, and A. Smola, "Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 953–964.

[167] A. Milisavljevic, F. Abate, T. Le Bras, B. Gosselin, M. Mancas, and K. Doré-Mazars, "Similarities and differences between eye and mouse dynamics during web pages exploration," *Frontiers in Psychology*, vol. 12, p. 554595, 2021.

[168] G. M. Grimes and J. S. Valacich, "Mind over mouse: The effect of cognitive load on mouse movement behavior," in *2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015*. Association for Information Systems, 2015.

[169] H. Rheem, V. Verma, and D. V. Becker, "Use of mouse-tracking method to measure cognitive load," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 62, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2018, pp. 1982–1986.

[170] E. J. Williams, "Experimental designs balanced for the estimation of residual effects of treatments," *Australian Journal of Chemistry*, vol. 2, no. 2, pp. 149–168, 1949.

[171] A. Zubiaga, "A longitudinal assessment of the persistence of Twitter datasets," *Journal of the Association for Information Science and Technology*, vol. 69, no. 8, pp. 974–984, 2018.

[172] History.com Editors, "Boston Marathon Bombing," History.com, 2014. [Online]. Available: https://www.history.com/topics/21st-century/boston-marathon-bombings

[173] ——, "ISIL stages series of terrorist attacks in Paris, culminating in massacre at Bataclan theater," History.com, 2019. [Online]. Available: https://www.history.com/this-day-in-history/paris-terrorist-attacks-bataclan

[174] K. A. Neuendorf, *The content analysis guidebook.* sage, 2017.

[175] M. L. McHugh, "Interrater reliability: The kappa statistic," *Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[176] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[177] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

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

[179] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[180] D. A. Salazar, J. I. Vélez, and J. C. Salazar, "Comparison between svm and logistic regression: Which one is better to discriminate?" *Revista Colombiana de Estadistica*, vol. 35, no. 2, pp. 223–237, 2012. [Online]. Available: https://www.proquest.com/scholarly-journals/comparison-between-svm-logistic-regression-which/docview/1677632752/se-2

[181] M. Somvanshi, P. Chavan, S. Tambade, and S. V. Shinde, "A review of machine learning techniques using decision tree and support vector machine," in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*, 2016, pp. 1–7.

[182] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[183] A. Ng and M. Jordan, "On discriminative vs. generative classifiers: A comparison

of logistic regression and naive bayes," *Advances in neural information processing systems*, vol. 14, 2001.

[184] C. Molnar, *Interpretable Machine Learning*, 2nd ed. Self-published, 2022. [Online]. Available: https://christophm.github.io/interpretable-ml-book

[185] D. Qiu and A. C. Tamhane, "A comparative study of the K-means algorithm and the normal mixture model for clustering: Univariate case," *Journal of Statistical Planning and Inference*, vol. 137, no. 11, pp. 3722–3740, 2007.

[186] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[187] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in Python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[188] M. Bramer, "Avoiding overfitting of decision trees," *Principles of data mining*, pp. 119–134, 2007.

[189] H. Wang and M. Song, "Ckmeans. 1d. dp: Optimal k-means clustering in one dimension by dynamic programming," *The R journal*, vol. 3, no. 2, p. 29, 2011.

[190] B. M. Velichkovsky, S. M. Dornhoefer, S. Pannasch, and P. J. Unema, "Visual fixations and level of attentional processing," in *Proceedings of the 2000 symposium on eye tracking research & applications*, 2000, pp. 79–85.

[191] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 50, no. 9. Sage publications Sage CA: Los Angeles, CA, 2006, pp. 904–908.

[192] J. Bruin. (2011, Feb.) newtest: Command to compute new test. [Online]. Available: https://stats.oarc.ucla.edu/stata/ado/analysis/

[193] A. J. Viera, J. M. Garrett *et al.*, "Understanding interobserver agreement: The kappa statistic," *Fam med*, vol. 37, no. 5, pp. 360–363, 2005.

[194] A. Medlar and D. Glowacka, "How Consistent is Relevance Feedback in Exploratory Search?" in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, ser. CIKM '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 1615–1618. [Online]. Available: https://doi.org/10.1145/3269206.3269297

[195] A. K. Dubey, K. Gulabani, A. Guwalani, and R. Rathi, "Pragmatic study on HCI Laws," *International Journal of Science, Engineering and Technology Research (IJSETR)*, vol. 3, no. 12, 2014.

[196] P. Nayak, "Understanding searches better than ever before," https://blog.google/products/search/search-language-understanding-bert/, 2019, accessed: 2022-08-14.

[197] S. Eliot and J. Rose, *A Companion to the History of the Book*. John Wiley & Sons, 2009, vol. 98.

[198] C. N. Mooers, *The theory of digital handling of non-numerical information and its implications to machine economics*. Zator Company, 1950, no. 48.

[199] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. USA: McGraw-Hill, Inc., 1986.

[200] M. Taube, C. D. Gull, and I. S. Wachtel, "Unit terms in coordinate indexing," *Journal of the American Society for Information Science*, vol. 3, no. 4, p. 213, 1952.

[201] H. P. Luhn, "The automatic creation of literature abstracts," *IBM Journal of research and development*, vol. 2, no. 2, pp. 159–165, 1958.

[202] I. Ruthven, "Interactive information retrieval," *Annual review of information science and technology*, vol. 42, pp. 43–92, 2008.

[203] B. Shneiderman, D. Byrd, and W. B. Croft, "Sorting out Searching: A User-Interface Framework for Text Searches," *Commun. ACM*, vol. 41, no. 4, p. 95–98, apr 1998. [Online]. Available: https://doi.org/10.1145/273035.273069

[204] M. J. Bates, "The design of browsing and berrypicking techniques for the online search interface," *Online review*, 1989.

[205] C. C. Kuhlthau, "Inside the search process: Information seeking from the user's perspective," *Journal of the American society for information science*, vol. 42, no. 5, pp. 361–371, 1991.

[206] P. Ingwersen, "Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory," *Journal of Documentation*, 1996.

[207] N. J. Belkin, "Anomalous states of knowledge as a basis for information retrieval," *Canadian journal of information science*, vol. 5, no. 1, pp. 133–143, 1980.

[208] R. W. White, B. Kules, S. M. Drucker *et al.*, "Supporting exploratory search, introduction, special issue, communications of the acm," *Communications of the ACM*, vol. 49, no. 4, pp. 36–39, 2006.

[209] C. Hendahewa and C. Shah, "Evaluating user search trails in exploratory search tasks," *Information Processing & Management*, vol. 53, no. 4, pp. 905–922, 2017.

[210] A. Soufan, I. Ruthven, and L. Azzopardi, "Searching the literature: An analysis of an exploratory search task," in *ACM SIGIR Conference on Human Information Interaction and Retrieval*, 2022, pp. 146–157.

[211] H. Cui, J.-R. Wen, J.-Y. Nie, and W.-Y. Ma, "Probabilistic query expansion using query logs," in *Proceedings of the 11th international conference on World Wide Web*, 2002, pp. 325–332.

[212] J. Rocchio, "Relevance feedback in information retrieval," *The Smart retrieval system-experiments in automatic document processing*, pp. 313–323, 1971.

[213] X. Shen, B. Tan, and C. Zhai, “Context-sensitive information retrieval using implicit feedback,” in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 2005, pp. 43–50.

[214] T. Joachims, “Optimizing search engines using clickthrough data,” in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2002, pp. 133–142.

[215] D. Kelly and J. Teevan, “Implicit feedback for inferring user preference: A bibliography,” in *SIGIR forum*, vol. 37, no. 2, 2003, pp. 18–28.

[216] R. W. White, J. M. Jose, C. J. van Rijsbergen, and I. Ruthven, “A simulated study of implicit feedback models,” in *European Conference on Information Retrieval.* Springer, 2004, pp. 311–326.

[217] K. Sugiyama, K. Hatano, and M. Yoshikawa, “Adaptive web search based on user profile constructed without any effort from users,” in *Proceedings of the 13th international conference on World Wide Web.* ACM, 2004, pp. 675–684.

[218] D. Kelly and N. J. Belkin, “Display time as implicit feedback: Understanding task effects,” in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 2004, pp. 377–384.

[219] I. Arapakis, K. Athanasakos, and J. M. Jose, “A comparison of general vs personalised affective models for the prediction of topical relevance,” in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.* ACM, 2010, pp. 371–378.

[220] S. B. Huffman and M. Hochster, “How well does result relevance predict session satisfaction?” in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval.* ACM, 2007, pp. 567–574.

[221] D. R. Swanson, “Subjective versus objective relevance in bibliographic retrieval systems,” *The Library Quarterly*, vol. 56, no. 4, pp. 389–398, 1986.

[222] P. Borlund, “The concept of relevance in IR,” *Journal of the American Society for information Science and Technology*, vol. 54, no. 10, pp. 913–925, 2003.

[223] T. Saracevic, “Relevance: A review of and a framework for the thinking on the notion in information science,” *Journal of the American Society for information science*, vol. 26, no. 6, pp. 321–343, 1975.

[224] Y. Moshfeghi, L. R. Pinto, F. E. Pollick, and J. M. Jose, “Understanding relevance: An fMRI study,” in *European conference on information retrieval.* Springer, 2013, pp. 14–25.

[225] C. F. Camerer, G. Loewenstein, and D. Prelec, “Neuroeconomics: Why economics needs brains,” *The Scandinavian Journal of Economics*, vol. 106, no. 3, pp. 555–579, 2004.

[226] N. Lee, A. J. Broderick, and L. Chamberlain, “What is ‘neuromarketing’? A discussion and agenda for future research,” *International journal of psychophysiology*, vol. 63, no. 2, pp. 199–204, 2007.

[227] A. Dimoka, P. A. Pavlou, and F. D. Davis, “Neuro-IS: The potential of cognitive neuroscience for information systems research,” in *Proceedings of the 28th international conference on information systems*, 2007, pp. 1–20.

[228] Y. Moshfeghi, “Neurasearch: Neuroscience and information retrieval,” in *CEUR Workshop Proceedings*, vol. 2950, 2021, pp. 193–194.

[229] M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski, “Predicting term-relevance from brain signals,” in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 425–434.

[230] M. Allegretti, Y. Moshfeghi, M. Hadjigeorgieva, F. E. Pollick, J. M. Jose, and G. Pasi, “When relevance judgement is happening? An EEG-based study,” in

Proceedings of the 38th international acm sigir conference on research and development in information retrieval, 2015, pp. 719–722.

[231] S. Thorpe, D. Fize, and C. Marlot, “Speed of processing in the human visual system,” *nature*, vol. 381, no. 6582, pp. 520–522, 1996.

[232] L. Dubreuil-Vall, P. Chau, G. Ruffini, A. S. Widge, and J. A. Camprodon, “tDCS to the left DLPFC modulates cognitive and physiological correlates of executive function in a state-dependent manner,” *Brain stimulation*, vol. 12, no. 6, pp. 1456–1463, 2019.

[233] M. J. Eugster, T. Ruotsalo, M. M. Spapé, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski, “Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals,” *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[234] R. J. Chabot, D. H. York, and W. A. Waugh, “The late positive component of the evoked waveform: Relationship to word recognition processes,” *International journal of neuroscience*, vol. 22, no. 3-4, pp. 299–313, 1984.

[235] H. H. Kim and Y. H. Kim, “ERP/MMR algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos,” *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, pp. 931–941, 2019.

[236] Z. Pinkosova, W. J. McGeown, and Y. Moshfeghi, “The Cortical Activity of Graded Relevance,” in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR ’20. New York, NY, USA: Association for Computing Machinery, 2020, p. 299–308. [Online]. Available: https://doi.org/10.1145/3397271.3401106

[237] Y. Moshfeghi, P. Triantafillou, and F. E. Pollick, “Understanding information need: An fMRI study,” in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 335–344.

[238] D. Michalkova, M. Parra-Rodriguez, and Y. Moshfeghi, "Information Need Awareness: An EEG study," in *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022, pp. 610–621.

[239] T. D. Wilson, "On user studies and information needs," *Journal of Documentation*, 1981.

[240] T. Saracevic, "Relevance reconsidered," in *Proceedings of the second conference on conceptions of library and information science (CoLIS 2)*, 1996, pp. 201–218.

[241] G. Jacucci, O. Barral, P. Daee, M. Wenzel, B. Serim, T. Ruotsalo, P. Pluchino, J. Freeman, L. Gamberini, S. Kaski *et al.*, "Integrating neurophysiologic relevance feedback in intent modeling for information retrieval," *Journal of the Association for Information Science and Technology*, vol. 70, no. 9, pp. 917–930, 2019.

[242] A. Diamond, "Executive functions," *Annual review of psychology*, vol. 64, p. 135, 2013.

[243] C. Rottschy, R. Langner, I. Dogan, K. Reetz, A. R. Laird, J. B. Schulz, P. T. Fox, and S. B. Eickhoff, "Modelling neural correlates of working memory: A coordinate-based meta-analysis," *Neuroimage*, vol. 60, no. 1, pp. 830–846, 2012.

[244] J. Duncan and A. M. Owen, "Common regions of the human frontal lobe recruited by diverse cognitive demands," *Trends in neurosciences*, vol. 23, no. 10, pp. 475–483, 2000.

[245] N. U. Dosenbach, D. A. Fair, A. L. Cohen, B. L. Schlaggar, and S. E. Petersen, "A dual-networks architecture of top-down control," *Trends in cognitive sciences*, vol. 12, no. 3, pp. 99–105, 2008.

[246] T. A. Niendam, A. R. Laird, K. L. Ray, Y. M. Dean, D. C. Glahn, and C. S. Carter, "Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 12, no. 2, pp. 241–268, 2012.

[247] G. Tabibnia, J. D. Creswell, T. E. Kraynak, C. Westbrook, E. Julson, and H. A. Tindle, "Common prefrontal regions activate during self-control of craving, emotion, and motor impulses in smokers," *Clinical Psychological Science*, vol. 2, no. 5, pp. 611–619, 2014.

[248] M. Neta, S. M. Nelson, and S. E. Petersen, "Dorsal anterior cingulate, medial superior frontal cortex, and anterior insula show performance reporting-related late task control signals," *Cerebral cortex*, vol. 27, no. 3, pp. 2154–2165, 2017.

[249] A. D. Baddeley and G. Hitch, "Working Memory," in *Psychology of Learning and Motivation*, G. H. Bower, Ed. Academic Press, 1974, vol. 8, pp. 47 – 89. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0079742108604521

[250] E. K. Hussey, N. Ward, K. Christianson, and A. F. Kramer, "Language and memory improvements following tDCS of left lateral prefrontal cortex," *PloS one*, vol. 10, no. 11, p. e0141417, 2015.

[251] A. Gazzaley and A. C. Nobre, "Top-down modulation: Bridging selective attention and working memory," *Trends in cognitive sciences*, vol. 16, no. 2, pp. 129–135, 2012.

[252] M. D'Esposito, B. R. Postle, and B. Rypma, "Prefrontal cortical contributions to working memory: Evidence from event-related fMRI studies," in *Executive control and the frontal lobe: Current issues.* Springer, 2000, pp. 3–11.

[253] T. Iuculano and R. C. Kadosh, "The mental cost of cognitive enhancement," *Journal of Neuroscience*, vol. 33, no. 10, pp. 4482–4486, 2013.

[254] A. K. Barbey, M. Koenigs, and J. Grafman, "Dorsolateral prefrontal contributions to human working memory," *cortex*, vol. 49, no. 5, pp. 1195–1205, 2013.

[255] R. Manenti, S. F. Cappa, P. M. Rossini, and C. Miniussi, "The role of the prefrontal cortex in sentence comprehension: An rTMS study," *cortex*, vol. 44, no. 3, pp. 337–344, 2008.

[256] E. K. Miller, "The prefontral cortex and cognitive control," *Nature reviews neuroscience*, vol. 1, no. 1, pp. 59–65, 2000.

[257] J. T. Nelson, R. A. McKinley, E. J. Golob, J. S. Warm, and R. Parasuraman, "Enhancing vigilance in operators with prefrontal cortex transcranial direct current stimulation (tDCS)," *Neuroimage*, vol. 85, pp. 909–917, 2014.

[258] P. A. Reuter-Lorenz, J. Jonides, E. E. Smith, A. Hartley, A. Miller, C. Marshuetz, and R. A. Koeppe, "Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET," *Journal of cognitive neuroscience*, vol. 12, no. 1, pp. 174–187, 2000.

[259] R. Cabeza, "Hemispheric asymmetry reduction in older adults: The HAROLD model." *Psychology and aging*, vol. 17, no. 1, p. 85, 2002.

[260] J. Klaus and D. J. Schutter, "The role of left dorsolateral prefrontal cortex in language processing," *Neuroscience*, vol. 377, pp. 197–205, 2018.

[261] M.-F. Marin, J. A. Camprodon, D. D. Dougherty, and M. R. Milad, "Device-based brain stimulation to augment fear extinction: Implications for PTSD treatment and beyond," *Depression and anxiety*, vol. 31, no. 4, pp. 269–278, 2014.

[262] A. Datta, V. Bansal, J. Diaz, J. Patel, D. Reato, and M. Bikson, "Gyri-precise head model of transcranial direct current stimulation: Improved spatial focality using a ring electrode versus conventional rectangular pad," *Brain stimulation*, vol. 2, no. 4, pp. 201–207, 2009.

[263] A. Thielscher, A. Antunes, and G. B. Saturnino, "Field modeling for transcranial magnetic stimulation: A useful tool to understand the physiological effects of TMS?" in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 222–225.

[264] N. Khadka, H. Borges, B. Paneri, T. Kaufman, E. Nassis, A. L. Zannou, Y. Shin, H. Choi, S. Kim, K. Lee *et al.*, "Adaptive current tDCS up to 4 mA," *Brain stimulation*, vol. 13, no. 1, pp. 69–79, 2020.

[265] M.-F. Kuo, W. Paulus, and M. A. Nitsche, "Therapeutic effects of non-invasive brain stimulation with direct currents (tDCS) in neuropsychiatric diseases," *Neuroimage*, vol. 85, pp. 948–960, 2014.

[266] L. Dumont, F. Larochelle-Brunet, H. Théoret, R. Riedl, S. Sénécal, and P.-M. Leger, "Non-invasive brain stimulation in information systems research: A proof-of-concept study," *PloS one*, vol. 13, no. 7, p. e0201128, 2018.

[267] G. Learmonth, C. S. Benwell, G. Märker, D. Dascalu, M. Checketts, C. Santosh, M. Barber, M. Walters, K. W. Muir, and M. Harvey, "Non-invasive brain stimulation in stroke patients (NIBS): A prospective randomized open blinded end-point (PROBE) feasibility trial using transcranial direct current stimulation (tDCS) in post-stroke hemispatial neglect," *Neuropsychological Rehabilitation*, vol. 31, no. 8, pp. 1163–1189, 2021.

[268] L. Xiang, W. Jing, W. Pu *et al.*, "Transcranial Direct Current Stimulation for Parkinson's Disease: A Systematic Review and Meta-Analysis," *Frontiers in Aging Neuroscience*, p. 691, 2021.

[269] M. G. Philiastides, R. Auksztulewicz, H. R. Heekeren, and F. Blankenburg, "Causal role of dorsolateral prefrontal cortex in human perceptual decision making," *Current biology*, vol. 21, no. 11, pp. 980–983, 2011.

[270] M. Daneman and P. M. Merikle, "Working memory and language comprehension: A meta-analysis," *Psychonomic bulletin & review*, vol. 3, no. 4, pp. 422–433, 1996.

[271] C. Cerruti and G. Schlaug, "Anodal transcranial direct current stimulation of the prefrontal cortex enhances complex verbal associative thought," *Journal of cognitive neuroscience*, vol. 21, no. 10, pp. 1980–1987, 2009.

[272] N. Metuki, T. Sela, and M. Lavidor, "Enhancing cognitive control components of insight problems solving by anodal tDCS of the left dorsolateral prefrontal cortex," *Brain Stimulation*, vol. 5, no. 2, pp. 110–115, 2012.

[273] C. S. Carver and T. L. White, “Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales.” *Journal of personality and social psychology*, vol. 67, no. 2, p. 319, 1994.

[274] J. A. Miler, D. Meron, D. S. Baldwin, and M. Garner, “The effect of prefrontal transcranial direct current stimulation on attention network function in healthy volunteers,” *Neuromodulation: Technology at the Neural Interface*, vol. 21, no. 4, pp. 355–361, 2018.

[275] J. Fan, B. D. McCandliss, T. Sommer, A. Raz, and M. I. Posner, “Testing the efficiency and independence of attentional networks,” *Journal of cognitive neuroscience*, vol. 14, no. 3, pp. 340–347, 2002.

[276] S. Arora, M. A. Lawrence, and R. M. Klein, “The attention network test database: ADHD and cross-cultural applications,” *Frontiers in Psychology*, vol. 11, p. 388, 2020.

[277] A. Johnstone and P. Marí-Beffa, “The effects of martial arts training on attentional networks in typical adults,” *Frontiers in psychology*, p. 80, 2018.

[278] T. S. Redick and R. W. Engle, “Working memory capacity and attention network test performance,” *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, vol. 20, no. 5, pp. 713–721, 2006.

[279] S. Rossi, M. Hallett, P. M. Rossini, A. Pascual-Leone, S. of TMS Consensus Group *et al.*, “Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research,” *Clinical neurophysiology*, vol. 120, no. 12, pp. 2008–2039, 2009.

[280] T. Minarik, B. Berger, L. Althaus, V. Bader, B. Biebl, F. Brotzeller, T. Fusban, J. Hegemann, L. Jesteadt, L. Kalweit *et al.*, “The importance of sample size for reproducibility of tDCS effects,” *Frontiers in Human Neuroscience*, vol. 10, p. 453, 2016.

[281] Q. Chi, X. Jiang, and Y. Lv, "Analysis on the Group Differences of Information Literacy among Local College Students," in *2021 4th International Conference on Information Systems and Computer Aided Education*, 2021, pp. 1893–1898.

[282] R. C. Oldfield *et al.*, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971.

[283] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing.* Association for Computational Linguistics, 11 2019. [Online]. Available: https://arxiv.org/abs/1908.10084

[284] D. McNeish, "Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction," *Multivariate Behavioral Research*, vol. 52, no. 5, pp. 661–670, 2017.

[285] V. A. Brown, "An introduction to linear mixed-effects modeling in R," *Advances in Methods and Practices in Psychological Science*, vol. 4, no. 1, p. 2515245920960351, 2021.

[286] M. Allen, D. Poggiali, K. Whitaker, T. R. Marshall, and R. A. Kievit, "Raincloud plots: A multi-platform tool for robust data visualization," *Wellcome open research*, vol. 4, 2019.

[287] L. Fassi and R. C. Kadosh, "Is it all in our head? When subjective beliefs about receiving an intervention are better predictors of experimental results than the intervention itself," *BioRxiv*, 2020.

[288] J. C. Horvath, O. Carter, and J. D. Forte, "Transcranial direct current stimulation: Five important issues we aren't discussing (but probably should be)," *Frontiers in systems neuroscience*, vol. 8, p. 2, 2014.

[289] A. Vergallito, S. Feroldi, A. Pisoni, and L. J. Romero Lauro, "Inter-Individual Variability in tDCS Effects: A Narrative Review on the Contribution of Stable, Variable, and Contextual Factors," *Brain Sciences*, vol. 12, no. 5, p. 522, 2022.

[290] F. Fregni, P. S. Boggio, M. Nitsche, F. Bermpohl, A. Antal, E. Feredoes, M. A. Marcolin, S. P. Rigonatti, M. T. Silva, W. Paulus *et al.*, “Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory,” *Experimental brain research*, vol. 166, no. 1, pp. 23–30, 2005.

[291] J. Fins, “Neuroethics and the lure of technology,” in *Oxford Handbook of Neuroethics*. Oxford University Press, 2012.

[292] N. Nozari, K. Woodard, and S. L. Thompson-Schill, “Consequences of cathodal stimulation for behavior: When does it help and when does it hurt performance?” *PloS one*, vol. 9, no. 1, p. e84338, 2014.

[293] A. Quartarone, F. Morgante, S. Bagnato, V. Rizzo, A. Sant’Angelo, E. Aiello, E. Reggio, F. Battaglia, C. Messina, and P. Girlanda, “Long lasting effects of transcranial direct current stimulation on motor imagery,” *Neuroreport*, vol. 15, no. 8, pp. 1287–1291, 2004.

[294] G. Learmonth, L. Buhôt, L. Möller, and R. Greinacher, “Ineffective single-blinding during 1mA transcranial direct current stimulation.” *Journal of Vision*, vol. 19, no. 10, pp. 277c–277c, 2019.

[295] C. D. Workman, A. C. Fietsam, and T. Rudroff, “Different effects of 2 mA and 4 mA transcranial direct current stimulation on muscle activity and torque in a maximal isokinetic fatigue task,” *Frontiers in human neuroscience*, vol. 14, p. 240, 2020.

[296] I. Papazova, W. Strube, B. Becker, B. Henning, T. Schwippel, A. J. Fallgatter, F. Padberg, U. Palm, P. Falkai, C. Plewnia *et al.*, “Improving working memory in schizophrenia: Effects of 1 mA and 2 mA transcranial direct current stimulation to the left DLPFC,” *Schizophrenia research*, vol. 202, pp. 203–209, 2018.

[297] S. Maccora, G. Giglia, N. Bolognini, G. Cosentino, M. Gangitano, G. Salemi, and F. Brighina, “Cathodal occipital tDCS is unable to modulate the sound induced flash illusion in migraine,” *Frontiers in human neuroscience*, vol. 13, p. 247, 2019.

[298] J. Sweller, J. J. Van Merrienboer, and F. G. Paas, “Cognitive architecture and instructional design,” *Educational psychology review*, vol. 10, no. 3, pp. 251–296, 1998.

[299] A. A. Benczúr, L. Kocsis, and R. Pálovics, “Online machine learning in big data streams,” *arXiv preprint arXiv:1802.05872*, 2018.

[300] E. Ophir, C. Nass, and A. D. Wagner, “Cognitive control in media multitaskers,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 37, pp. 15 583–15 587, 2009.

[301] M. Dubiel, “Exploring the impact of conversational strategies on user search experience in goal-oriented tasks in a voice-only domain,” Ph.D. dissertation, University of Strathclyde, 2021.

[302] F. Petroni, T. Rocktäschel, P. Lewis, A. Bakhtin, Y. Wu, A. H. Miller, and S. Riedel, “Language models as knowledge bases?” *arXiv preprint arXiv:1909.01066*, 2019.

[303] N. Ranasinghe, P. Jain, S. Karwita, D. Tolley, and E. Y.-L. Do, “Ambiotherm: Enhancing sense of presence in virtual reality by simulating real-world environmental conditions,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 1731–1742.

Appendix A

Twitter study

A.1 Information sheet

Research Title: Examining Information Overload in Social Networks

Researcher: Mohamed Amine Belabbes mohamed.belabbes@strath.ac.uk
Supervisor 1: Dr. Yashar Moshfeghi yashar.moshfeghi@strath.ac.uk
Supervisor 2: Prof. Ian Ruthven ian.ruthven@strath.ac.uk

Name of the department: Computer and Information Sciences
Ethics Approval No.: 1477

You are being invited to participate in a study involving the use of a replica of the social network Twitter. It is very important that you understand what these terms mean and what being involved in the study will mean for you. Please take time to read this information sheet and do not hesitate to contact us if you have any queries. Please remember that you can withdraw from the study at any time without giving a reason. You can take as much time as you need before deciding as to whether continue with the study.

Why are we doing this?

The aim of the present study is to investigate information overload in social networks. The experiment is designed to create the situation of experiencing information overload. We want to investigate whether it is possible to reproduce information overload under an experimental setting. To achieve this, you will be using a minimalistic clone of Twitter.

Who can take part in this study?

All participants are required to be healthy and fit the 18-40 years old age group, participants should also be fluent in English.

Do I have to take part?

Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any question, without providing a reason.

What is Twitter ?

In this experiment, we will be using a minimalistic version Twitter which is a social network where anyone can post about anything and are only limited by the length of the posts (tweets) which is 240 characters. In this experiment, you won't be asked to tweet but rather to follow the development in your timeline.

What will happen to me if I take part?

If you agree to take part in the experiment you will be given a consent form to prove that you are happy to participate.

The experiment will take place online over two sessions and it will be executed through Zoom. Once you have given your consent, you will have time to familiarise yourself with the Twitter clone. You will have the opportunity to ask any questions you may have. Once you feel comfortable, we will start. At the end of the study, you will be given a 7£ Amazon voucher.

How long does the study take?

Participation in this study will take around 30 minutes

What will you do in the project?

First you will have to complete an entry questionnaire where we will collect demographics data about you, subsequently you will have to fill in a pre-task questionnaire. You will then be given time to familiarise yourself with the Twitter clone.

In this experiment design, we will be using Zoom to simulate the experience of being part of a newsroom. You will have to narrate out loud the events as they are arriving on Twitter and whenever you feel overloaded say it loud. At the end, you will have to write a summary of the event which you were witnessing on Twitter. The scenario is as follows:

"You are a journalist and are part of a newsroom in charge of following the latest events through Twitter. During the next 10 minutes, you must follow the evolution of an event and keep your colleagues in the room updated. It means that you will have to narrate the development as they are happening on social media, and you will be required to share if you feel overloaded by saying it out loud. At the end of the 10 minutes, you will be writing a summary of what happened for your colleagues, it will need to contain key details about the event."

At the end of the 10 minutes, you will have to summarise the key aspects of the events such as number of causalities, location…etc. Your summary needs to be informative so your colleagues can understand what happened, who was impacted…etc

The experiment will conclude with an exit questionnaire and semi-structured interviews where you will have the opportunity to share your experience.

Will my taking part in this study be kept confidential?

Yes, all data collected from you will be treated confidentially, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from you.

The University of Strathclyde is committed to transparency and to complying with its responsibilities under data protection legislation. All collected data will be processed in accordance with the General Data Protection Regulation and the Data Protection Act 2018 and treated with the strict adherence to the Code of Practice of the University of Strathclyde. All personal and demographic data obtained will be used and presented in the aggregated format. Due to the sensitive nature of this research, data obtained in this experiment will not be openly available.

You may request for your personal data to be destroyed at any point in the future. Please note, however, that it will not be possible to remove your experimental data from analyses that have already been completed (e.g., once your data is combined with data from other participants). If you wish to request withdrawal of your data, please contact yashar.moshfeghi@strath.ac.uk.

Please also read our Privacy Notice for Research Participants (see supplement). This can also be found here: http://shorturl.at/jEJU6

What are the possible benefits?

If you take part to this study, you will be contributing in allowing the scientific community in better tackling the problem information overload in the context of social networks.

Who will have access to the information?

Only below-mentioned researchers (Dr Yashar Moshfeghi, Prof Ian Ruthven, Mohamed Amine Belabbes) will have access to the data. It is possible that the data may be used by the below- mentioned researchers for other similar ethically approved research protocols, where the same standards of confidentiality will apply. Due to the sensitive nature of the data, the data will not be shared (unless approved by the Principal Investigator, Dr Yashar Moshfeghi).

Where will the information be stored and how long will it be kept for?

The collected data will be stored and kept for as long as it is required by involved researchers. After that, the data will be securely deleted. The collected data will be stored privately at a secured location, which will be password protected. All data will be anonymised to the best possibilities.
The data will not be shared due to the sensitive nature (unless approved by the Principal Investigator, Dr Yashar Moshfeghi).

What will happen to the results of this study?

The results of this study may appear in a published article, where anonymity of participants will be preserved.

Who has reviewed the study?

The study has been reviewed by the University of Strathclyde, Department of Computer and Information Sciences Ethics committee.

If there are any further questions regarding this study, please contact any one of the study team:

Mohamed Amine Belabbes mohamed.belabbes@strath.ac.uk
Doctor Yashar Moshfeghi yashar.moshfeghi@strath.ac.uk
Professor Ian Ruthven ian.ruthven@strath.ac.uk

This research was granted ethical approval by the Department of Computer and Information Sciences Ethics Committee under application number 1477. If you have any questions or concerns, before, during or after the investigation, or wish to contact an independent person to whom any questions may be directed or further information may be sought form, contact details are provided below:

Department of Computer and Information Sciences Ethics Committee University of Strathclyde
Livingstone Tower
16 Richmond Street
Glasgow
G1 1XQ
United Kingdom
E-mail: ethics@cis.strath.ac.uk
Telephone: 0141 548 3707

A.2 Consent form

Research Informed Consent Form

Title of the project:
Examining Information Overload in Social Networks
Ethics approval no.: 1477
Researcher's name: Mohamed Amine Belabbes
Researcher's e-mail: mohamed.belabbes@strath.ac.uk
Name of department: Computer and Information Sciences

Please, read the following statements and insert your initials for each statement you agree with:

- I confirm that I have read and understood the information sheet for the experiment. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

  Participant's Initials: ....

- I understand that my participation is entirely voluntary and that I am free to withdraw from the experiment at any time, up to the point of completion, without having to give a reason and without any consequences.

  Participant's Initials: ....

- I understand that I can withdraw from the study any personal data (i.e. data which may identify me personally) at any time.

  Participant's Initials: ....

- I understand that anonymised data (i.e. data which does not identify me personally) cannot be withdrawn once they have been included in the study.

  Participant's Initials: ....

- I understand that any information recorded in the study will be treated confidentially and no information that identifies me will be made publicly available.

  Participant's Initials: ....

If you would like a copy of this consent form to keep, please ask the researcher.
If you have any complaints or concerns about this research, you can direct these to Departmental Ethics Committee, in writing by e-mail at: ethics@cis.strath.ac.uk

- I confirm that I have read and understood the Privacy Notice for Participants in Research Projects and understand how my personal information will be used and what will happen to it.

Participant's Initials: ....

- I understand that anonymised data will be stored in a secured location for as long as it will be required by involved researchers.

Participant's Initials: ....

- I consent to be a participant in this study.

Participant's Initials: ....

Dated signature of the investigator

Dated signature of the participant

If there are any further questions regarding this study, please contact any one of the study team:

Mohamed Amine Belabbes mohamed.belabbes@strath.ac.uk
Doctor Yashar Moshfeghi yashar.moshfeghi@strath.ac.uk
Professor Ian Ruthven ian.ruthven@strath.ac.uk

A.3 Debriefing form

Debriefing Form

University of Strathclyde Science

Title of the project:
Examining Information Overload in Social Networks
Ethics approval no.: 1477
Researcher's name: Mohamed Amine Belabbes
Researcher's e-mail: mohamed.belabbes@strath.ac.uk
Name of department: Computer and Information Sciences

Thank you for taking part in this research. The aim of this research is to investigate information overload in social networks.

If you would like more information about this study, once it is completed, please contact the researcher:
Mohamed Amine Belabbes mohamed.belabbes@strath.ac.uk
Doctor Yashar Moshfeghi yashar.moshfeghi@strath.ac.uk
Professor Ian Ruthven ian.ruthven@strath.ac.uk

If this research has caused you any distress or discomfort and you would like to speak to someone, please, contact the following sources of support and advice: The Disability & Wellbeing Service, University of Strathclyde, e-mail: disability-wellbeing@strath.ac.uk, tel: 0141 548 3402.

More information can be found at:
https://www.strath.ac.uk/professionalservices/studentcounselling/

A.4 Entry questionnaire

Entry questionnaire

Start of Block: User and task ID

*

Please input your user ID

____

____

____

____

____

JS

Select the date

Month (1)	▼ January (1) ... (150)
Day (2)	▼ January (1) ... (150)
Year (3)	▼ January (1) ... (150)

End of Block: User and task ID

Start of Block: Demographics

What is the highest level of school you have completed or the highest degree you have received?

- ◯ Some college but no degree (3)
- ◯ Associate degree in college (2-year) (4)
- ◯ Bachelor's degree in college (4-year) (5)
- ◯ Master's degree (6)
- ◯ Doctoral degree (7)

Please tick the gender that best describes you

- ◯ Male (1)
- ◯ Female (2)
- ◯ Prefer to self-describe: (3) ____
- ◯ Prefer not to say (4)

End of Block: Demographics

Start of Block: SN proficiency

Which social network have you used in the past? (You can select more than one)

- ☐ Facebook (1)
- ☐ Twitter (2)
- ☐ Instagram (3)
- ☐ Pinterest (4)
- ☐ Tumblr (5)

How often do you use social networks?

- ○ Daily (1)
- ○ 4-6 times a week (2)
- ○ 2-3 times a week (3)
- ○ Once a week (4)
- ○ Never (5)

How often do you use a Twitter?

- ○ Daily (1)
- ○ 4-6 times a week (2)
- ○ 2-3 times a week (3)
- ○ Once a week (4)
- ○ About once or twice a month (5)
- ○ Never (6)

End of Block: SN proficiency

Start of Block: Block 3

Instructions to candidate:

In the next 10 minutes, you will have to narrate an event as the information comes through Twitter. You will have to also say out loud if you feel overloaded or if you feel like there is too much information coming through. “You are a journalist and are part of a newsroom in charge of following the latest events through Twitter. During the next 10 minutes, you must follow the evolution of an event and keep your colleagues in the room updated. **It means that you will have to narrate the development as they are happening on social media, and you will be required to share if you feel overloaded by saying it out loud. You should LIKE (Click on the heart) tweets which you find are useful to succeed at the task.** At the end of the 10 minutes, you will be writing a summary of what happened for your colleagues, it will need to contain key details about the event.”

You may use the following questions as a guidance:
• What is the on-going event?
• Are there casualties?
• How many bombs were detonated? What were their locations?
• Were there additional undetonated bombs? If so, how many and where were they located?
• Was there an impact of the attack on the stock market? If so, how did the market react to it?
• How many people passed away? How many were injured?
• Is there information on the citizenship of the people involved? If so, where were they from?
• Is there CCTV footage available showing the attacker(s)?
• Were they requiring blood donations? If so, can you name centres collection blood or their location?
• Was the attack claimed by a specific group or nation?

End of Block: Block 3

Start of Block: Perceptions

The task we asked you to perform is going to be easy

○ Strongly agree (1)

○ Somewhat agree (2)

○ Neither agree nor disagree (3)

○ Somewhat disagree (4)

○ Strongly disagree (5)

The task we asked you to perform is going to be stressful

- Strongly agree (1)
- Somewhat agree (2)
- Neither agree nor disagree (3)
- Somewhat disagree (4)
- Strongly disagree (5)

The task we asked you to perform is interesting

- Strongly agree (1)
- Somewhat agree (2)
- Neither agree nor disagree (3)
- Somewhat disagree (4)
- Strongly disagree (5)

The task we asked you to perform is clear

- Strongly agree (1)
- Somewhat agree (2)
- Neither agree nor disagree (3)
- Somewhat disagree (4)
- Strongly disagree (5)

The task we asked you to perform is familiar

○ Strongly agree (1)

○ Somewhat agree (2)

○ Neither agree nor disagree (3)

○ Somewhat disagree (4)

○ Strongly disagree (5)

End of Block: Perceptions

A.5 Pre-task questionnaire

Pre-task questionnaire

Start of Block: User and topic ID

Please input your user ID

____

Event ID

◯ 1 (1)

◯ 2 (2)

End of Block: User and topic ID

Start of Block: Familiarity

How familiar are you with the event you are asked to follow?

◯ Extremely familiar (1)

◯ Moderately familiar (2)

◯ Somewhat familiar (3)

◯ Slightly familiar (4)

◯ Not at all familiar (5)

End of Block: Familiarity

A.6 Post-task questionnaire

Post-task questionnaire

Start of Block: Default Question Block

Please input your user ID

____

Event ID

◯ 1 (1)

◯ 2 (2)

End of Block: Default Question Block

Start of Block: Event summary

Please provide a summary of the event and provide as many details as possible (in any format you want)

____

____

____

____

____

End of Block: Event summary

Start of Block: Causes

For each statement, share your agreement level

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
"There was too much information" (1)	○	○	○	○	○
"There was too much duplicate information" (2)	○	○	○	○	○
"There was too much new information" (3)	○	○	○	○	○
"I wasn't sure about what to look for" (6)	○	○	○	○	○
"Twitter's interface didn't help to properly complete the task" (7)	○	○	○	○	○
"The experiment's environment was not good" (8)	○	○	○	○	○

End of Block: Causes

Start of Block: Manifestations

For each statement, share your agreement level

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
"I felt fatigue" (1)	○	○	○	○	○
"I felt anxious" (2)	○	○	○	○	○
"I felt sad" (3)	○	○	○	○	○
"I felt stressed" (6)	○	○	○	○	○
"I wasn't able to understand the whole story" (7)	○	○	○	○	○
"I wasn't able to properly focus" (8)	○	○	○	○	○

End of Block: Manifestations

Start of Block: NASA-TLX

Please complete the following:

Very low Very high

How mentally demanding was the task? ()	
How hurried/rushed was the pace of the task? ()	
How successful were you in accomplishing what you were asked to do? ()	
How hard did you have to work to accomplish your level of performance? ()	
How insecure were you? ()	
How discouraged were you? ()	
How irritated were you? ()	
How stressed were you? ()	
How annoyed were you? ()	

End of Block: NASA-TLX

A.7 Exit questionnaire

Exit questionnaire

Start of Block: User ID

Please input your user ID

____

End of Block: User ID

Start of Block: Experience

How similar did you feel our system was to the actual Twitter interface?

- ○ Extremely similar (1)
- ○ Similar (2)
- ○ Slightly similar (3)
- ○ Dissimilar (4)
- ○ Extremely dissimilar (5)

Which event was easier to narrate?

- ○ Boston Marathon attack (1)
- ○ Paris attack (2)

Why was it easier to narrate?

____

____

____

Do you have any general feedback you would like to give?

End of Block: Experience

Appendix B

Neuromodulation study

B.1 Information sheet

Research Title: The ability of tDCS to stimulate Working Memory and affect the interaction of users on search systems

Researcher: Mohamed Amine Belabbes mohamed.belabbes@strath.ac.uk
Supervisor: Frank Pollick Frank.Pollick@glasgow.ac.uk

You are being invited to participate in a study involving the use of a search system and transcranial direct current stimulation (tDCS). It is very important that you understand what these terms mean and what being involved in the study will mean for you. Please take time to read this information sheet and do not hesitate to contact us if you have any queries. Please remember that you can withdraw from the study at any time without giving a reason. You can take as much time as you need before making a decision as to whether or not to continue with the study.

Why are we doing this?

The main objective of the experiment is to investigate the effect of tDCS on the interaction of users on search system and better understand the role of working memory (WM). Evidence suggests that transcranial Direct Current Stimulation (tDCS) on the left dorsolateral prefrontal cortex, usually associated with WM, is able to improve performance. The aim of this study is to examine how tDCS stimulation impact the user interaction. Examining the effects of tDCS on WM has the potential to help us better unstand the important of WM in the performance of an information retrieval task.

Who can take part in this study?

In order to take part in this study you must be above 18 years of age and have completed and passed the tDCS safety screening. Also you are not eligible to participate in this study if you are:

* Pregnant / Elderly
* Have any metal in your body, including the eyes
* If you or a family member has history of epileptic seizures, strokes, photosensitivity or fall into one of the following categories:

i. Inpatient care
ii. Incapacity
iii. Chronic serious health conditions

iv. Permanent or long term conditions
v. Conditions requiring multiple treatments

* Have a pacemaker or other implantable device
* Have history of neuropsychiatric or neurological illness
* Suffer from visual disorders
* Are sleep deprived
* Under the influence of alcohol
* Have suffered concussion or traumatic brain injury at any point
* Are prone to spells of dizziness/labyrinthitis
* Suffer from a heart condition or other serious medical conditions.
* Have eaten food 2 to 3 hours prior to the experiment.
* Have taken recreational drugs in the past week.
* Are taking or have taken anti-malarial medications at the moment or within the last 3 days.
* Are an HGV driver.

If you are on prescription medications please tell us about these, as you may not be able to take part depending on the medication you are on. If you do not want to do this, then please do not take part in the study.

Do I have to take part?

Your participation is voluntary, and you are free to withdraw at any time, without giving any reason, and you are free to omit answering any particular question, without providing a reason.

What is tDCS?

tDCS stands for transcranial direct current stimulation. tDCS uses 2 large electrodes placed onto the subject's scalp with some conducting solution. These are then held in place with two elastic bands. A very low current is passed through these electrodes, which will be up to a maximum of 2 milliamp. The stimulation will last for a maximum of 20 minutes. For most people tDCS is a completely painless procedure, but some people do feel a slight tingling sensation under the electrodes, especially when the current is switched on. Subjects usually describe this as being similar to an itching sensation. The effects will be minimised by increasing the current very slowly initially, which usually stops this tingling, but remember you can always ask the researcher to stop the stimulation at any point if you become uncomfortable.

What is Moogle ?

The Moogle search system is a search system developed to simulate a real life interaction with a search system.

What will happen to me if I take part?

If you agree to take part in the experiment you will be given a tDCS safety questionnaire. This information will only be used for exclusion from the study for the safety of a possible participant. You will be required to fill out these forms and sign a consent form before the experiment proceeds.

The testing will take place at the School of Psychology, 58 Hillhead Street, Glasgow. When you arrive a researcher will be there to meet you and show you the room where the experiment will take place. The researcher will go through the safety questionnaire with you and ask you sign a consent form. Then you will have time to familiarise yourself with the Moogle search system and the tDCS. You will have the opportunity to ask any questions you may have. You will be fitted with tDCS. You will be asked to sit in a chair and perform search tasks. You are able to stop participating at any time without providing any advanced notice or an explanation.

N.B. If you experience any of the following symptoms during this study, remove the head mounted display immediately and inform the researcher: (1) altered vision; (2) lightheadedness; (3) dizziness; (4) involuntary movements such as eye or muscle twitching; (5) confusion; (6) nausea; (7) loss of awareness; (8) convulsions; (9) cramps; (10) disorientation; (11) motion sickness; (12) Discomfort or pain in head or eyes; (13) Any other symptoms that you would class as atypical or unusual.

How long does the study take?

Participation in this study will take around 1 hour and 30 min.

What are the side effects of taking part?

tDCS uses a very low current and is not known to be harmful. There have been many studies throughout the world using this technique and no side effects have been reported, apart from the slight tingling feeling mentioned above, and occasional headaches. However, as with all techniques that directly stimulate the brain, tDCS has the possibility to induce seizures in people who are more sensitive to them. In order to find out whether you are likely to be more sensitive to seizures we ask you to fill in a safety questionnaire. Although no-one has had a seizure with the technique, it is very important you fill in this questionnaire accurately and if you are at an increased risk of seizures you will not be able to continue with the study. If you wish to read any literature on this, we would be happy to provide it for you.

Will my taking part in this study be kept confidential?

Yes, all data collected from you will be treated confidentially, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from you.

What are the possible benefits?

If you take part to this study you will have a chance to experience a non-invasive brain stimulation (NIBS) technique. It was shown to improve performance on various tasks. In addition, you will be part of the first cohort using tDCS on an information retrieval task.

What will happen to the results of this study?

The results of this study may appear in a published article, where anonymity of participants will be preserved.

Who has reviewed the study?

The study has been reviewed by the University of Glasgow, College of Science and Engineering Ethics Committee.

B.2 Safety questionnaire

Code Participant: __ __ | Date: __ __ / __ __ / __ __ (Day / Month / Year)

Investigator name: ____

Questionnaire tDCS

Have you ever:

▪ Had tDCS before?	☐ Yes	☐ No
▪ Had an adverse reaction to tDCS?	☐ Yes	☐ No
▪ Had a seizure?	☐ Yes	☐ No
▪ Had an unexplained spell of loss of consciousness?	☐ Yes	☐ No
▪ Had any brain-related, neurological injury or illnesses?	☐ Yes	☐ No
▪ Do you have any metal in your head (outside the mouth) such as shrapnel, surgical clips, cochlear implant or fragments from welding?	☐ Yes	☐ No
▪ Do you have any implanted medical devices such as cardiac pacemakers, or medical pumps?	☐ Yes	☐ No
▪ Do you suffer from frequent or severe headaches?	☐ Yes	☐ No
▪ Are you taking any medications (see list below)*?	☐ Yes	☐ No
▪ Have you recently taken any recreational drug or alcohol?	☐ Yes	☐ No
▪ Are you sleep deprived?	☐ Yes	☐ No
▪ Are you pregnant, or are you sexually active and not sure whether you might be pregnant?	☐ Yes	☐ No
▪ Does anyone in your family have epilepsy?	☐ Yes	☐ No
▪ Do you need any further explanation of tDCS or its associated risks?	☐ Yes	☐ No

FOR ANY « YES » RESPONSE, PLEASE PROVIDE DETAILED INFORMATION

____

____

____

SIGNATURES

Participant: ______________________________ Date: __ __ / __ __ / __ __

Investigator: ______________________________ Date: __ __ / __ __ / __ __ (Day / Month / Year)

***Excluded medications (as per Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of TMS Consensus Group. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clinical neurophysiology, 120(12), 2008-2039).**

Imipramine, amitriptyline, doxepine, nortriptyline, maprotiline, chlorpromazine, clozapine, foscarnet, ganciclovir, ritonavir, amphetamines, cocaine, (MDMA, ecstasy), phencyclidine (PCP, angel's dust), ketamine, gamma-hydroxybutyrate (GHB), alcohol, theophylline, mianserin, fluoxetine, fluvoxamine, paroxetine, sertraline, citalopram, reboxetine, venlafaxine, duloxetine, bupropion, mirtazapine, fluphenazine, pimozide, haloperidol, olanzapine, quetiapine, aripiprazole, ziprasidone, risperidone, chloroquine, mefloquine, imipenem, penicillin, ampicillin, cephalosporins, metronidazole, isoniazid, levofloxacin, cyclosporin, chlorambucil, vincristine, methotrexate, cytosine arabinoside, BCNU, lithium, anticholinergics, antihistamines, sympathomimetics. Withdrawal from alcohol, barbiturates, benzodiazepines, meprobamate, chloral hydrate.

B.3 Consent form

Consent Form

Title of Experiment: The ability of tDCS to stimulate Working Memory and affect the performance of users on search systems

Name of Researcher: Mohamed Amine Belabbes

1. I confirm that I have read and understand the Information sheet for the above study and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and am free to omit answering any particular question, without providing a reason.

3. I give consent for my actions to be recorded during the study.

4. I understand that all data collected from me will be treated confidentially, will be seen in its raw form only by the experimenters, and if published will not be identifiable as coming from me.

5. I understand that I should not log in to, or use, any personal web-based services on supplied devices during the course of the experiment, as all usage is being recorded and this could compromise my privacy.

6. I agree to take part in the above study.

Name of Participant	**Date**	**Signature**
Researcher	**Date**	**Signature**

Frank Pollick
Frank.Pollick@glasgow.ac.uk
Supervisor

If there are any further questions regarding this study, please contact any one of the study team:

Mohamed Amine Belabbes mohamed.belabbes@strath.ac.uk
Professor Frank Polick Frank.Pollick@glasgow.ac.uk

This study adheres to the BPS ethical guidelines, and has been approved by the Ethics Committee of the Glasgow University College of Science and Engineering

B.4 Entry questionnaire

Entry questionnaire

Start of Block: User and task ID

Please input your user ID

____

Session number

○ 1

○ 2

Enter a date:

____

End of Block: User and task ID

Start of Block: Demographics

What is your country of birth?

▼ Afghanistan ... Zimbabwe

What is your year of birth?

____

Page 1 of 7

What is the highest level of school you have completed or the highest degree you have received?

- ◯ Less than high school degree
- ◯ High school graduate (high school diploma or equivalent including GED)
- ◯ Some college but no degree
- ◯ Associate degree in college (2-year)
- ◯ Bachelor's degree in college (4-year)
- ◯ Master's degree
- ◯ Doctoral degree
- ◯ Professional degree (JD, MD)

Please tick the gender that best describes you

- ◯ Male
- ◯ Female
- ◯ Prefer to self-describe: ______________________________
- ◯ Prefer not to say

What is your english language proficiency?

- ◯ Fluent
- ◯ Intermediate
- ◯ Basic

End of Block: Demographics

Start of Block: Search engines proficiency

How would you describe your level of proficiency with search engines?

- ○ Excellent
- ○ Good
- ○ Average
- ○ Poor
- ○ Terrible

Which search engines have you used in the past? (You can select more than one)

- ▢ Google
- ▢ Bing
- ▢ Yahoo
- ▢ DuckDuckgo
- ▢ Startpage

How often do you use a search engine?

- ◯ Daily
- ◯ 4-6 times a week
- ◯ 2-3 times a week
- ◯ Once a week
- ◯ Never

End of Block: Search engines proficiency

Start of Block: Perceptions

How difficult do you think this experiment will be?

- ◯ Extremely easy
- ◯ Somewhat easy
- ◯ Neither easy nor difficult
- ◯ Somewhat difficult
- ◯ Extremely difficult

End of Block: Perceptions

Start of Block: EHQ

In the next questions, you will have to select your strongest hand for each category

Writing

- [ ] Left
- [ ] Right

Drawing

- [ ] Left
- [ ] Right

Throwing

- [ ] Left
- [ ] Right

Scissors

- [ ] Left
- [ ] Right

Toothbrush

- [ ] Left
- [ ] Right

Knife (without fork)

- [ ] Left
- [ ] Right

Spoon

- [ ] Left
- [ ] Right

Broom (upper hand)

- [ ] Left
- [ ] Right

Striking Match (match)

- [ ] Left
- [ ] Right

Opening box (lid)

- [ ] Left
- [ ] Right

Which foot do you prefer to kick with ?

- [ ] Left
- [ ] Right

Which eye do you use when using only one ?

- [ ] Left
- [ ] Right

End of Block: EHQ

Start of Block: Thank you !

Thank you very much !

End of Block: Thank you !

B.5 Pre-task questionnaire

Pre-task questionnaire

Start of Block: User and task ID

Please input your user ID

____

Please input the task ID

____

Session number

○ 1

○ 2

End of Block: User and task ID

Start of Block: Topics choice

*

Page 1 of 4

Please tick the topics' number you have chosen:

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7

Why did you choose these topics?

____

____

____

____

____

End of Block: Topics choice

Start of Block: Perceptions

The task we asked you to perform is going to be easy

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

The task we asked you to perform is going to be stressful

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

The task we asked you to perform is interesting

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

The task we asked you to perform is clear

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

The task we asked you to perform is familiar

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

End of Block: Perceptions

Start of Block: Block 3

Thank you ! The next step is performing the task using Moogle, at the end of the task you will be asked to give an answer to the task given.

End of Block: Block 3

B.6 Post-task questionnaire

Post-task questionnaire

Start of Block: Default Question Block

Please input your user ID

____

Please input the task ID

____

Please input the topic ID

____

Session number

○ 1

○ 2

End of Block: Default Question Block

Start of Block: Block 5

Please provide your answer to the given topic

____

____

____

____

End of Block: Block 5

Start of Block: Knowledge

How much more do you know about the topic assigned?

○ A great deal

○ A lot

○ A moderate amount

○ A little

○ None at all

End of Block: Knowledge

Start of Block: Block 5

How was your experience with Moogle during this task?

○ Extremely satisfied

○ Somewhat satisfied

○ Neither satisfied nor dissatisfied

○ Somewhat dissatisfied

○ Extremely dissatisfied

End of Block: Block 5

Start of Block: Perception

The task we asked you to perform was easy

- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Neither agree nor disagree
- ◯ Somewhat disagree
- ◯ Strongly disagree

The task we asked you to perform was stressful

- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Neither agree nor disagree
- ◯ Somewhat disagree
- ◯ Strongly disagree

The task we asked you to perform was interesting

- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Neither agree nor disagree
- ◯ Somewhat disagree
- ◯ Strongly disagree

The task we asked you to perform was familiar

- ◯ Strongly agree
- ◯ Somewhat agree
- ◯ Neither agree nor disagree
- ◯ Somewhat disagree
- ◯ Strongly disagree

End of Block: Perception

Start of Block: Thank you

Thank you very much !

End of Block: Thank you

B.7 Exit questionnaire

Exit questionnaire

Start of Block: User ID

Please input your user ID

____

Session Number

○ 1 (1)

○ 2 (2)

End of Block: User ID

Start of Block: Block 4

Did you experience any of these side-effects during the session?

	Not all (1)	Slightly (2)	Moderatly (3)	Very (4)	Extremely (5)
Headaches (28)	○	○	○	○	○
Tingling (29)	○	○	○	○	○
Itching (30)	○	○	○	○	○
Burning (31)	○	○	○	○	○
Pain (32)	○	○	○	○	○

If yes, how long did these side effects last? (click only if relevant)

	Not applicable (4)	A few seconds (5)	1-2 mins (6)	3-5 mins (7)	6-10 mins (8)	11-15 mins (9)
Headache (1)	○	○	○	○	○	○
Tingling (2)	○	○	○	○	○	○
Itching (3)	○	○	○	○	○	○
Burning (4)	○	○	○	○	○	○
Pain (5)	○	○	○	○	○	○

Display This Question:

If Session Number = 2

Can you guess which of the two sessions involved inactive ("sham") stimulation?

○ Session 1 (1)

○ Session 2 (2)

Display This Question:

If Session Number = 2

How sure are you?

- 1 (1)
- 2 (2)
- 3 (3)
- 4 (11)
- 5 (12)
- 6 (13)
- 7 (14)
- 8 (15)
- 9 (16)
- 10 (17)

Display This Question:

If Session Number = 2

Can you explain your reasoning?

____

End of Block: Block 4

Start of Block: General feeling

How much more do you feel you know about the researched topics?

- ◯ A great deal (1)
- ◯ A lot (2)
- ◯ A moderate amount (3)
- ◯ A little (4)
- ◯ None at all (5)

How difficult were the tasks?

- ◯ Extremely easy (1)
- ◯ Moderately easy (2)
- ◯ Slightly easy (3)
- ◯ Neither easy nor difficult (4)
- ◯ Slightly difficult (5)

End of Block: General feeling

Start of Block: Experience

How was your experience with Moogle?

- ◯ Extremely satisfied (1)
- ◯ Somewhat satisfied (2)
- ◯ Neither satisfied nor dissatisfied (3)
- ◯ Somewhat dissatisfied (4)
- ◯ Extremely dissatisfied (5)

Which topic did you prefer?

____

Why did you prefer that task?

____

____

____

____

____

Did you feel any discomfort during the experiment? Please describe

____

Do you have any general feedback you would like to give?

____

End of Block: Experience

Start of Block: BIS/BAS

Each item of this questionnaire is a statement that a person may either agree with or disagree with. For each item, indicate how much you agree or disagree with what the item says. Please respond to all the items; do not leave any blank. Choose only one response to each statement. Please be as accurate and honest as you can be. Respond to each item as if it were the only item. That is, don't worry about being "consistent" in your responses

A person's family is the most important thing in life

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

Even if something bad is about to happen to me, I rarely experience fear or nervousness

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

I go out of my way to get things I want

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

When I'm doing well at something I love to keep at it

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

I'm always willing to try something new if I think it will be fun

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

How I dress is important to me

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

When I get something I want, I feel excited and energized

- Very true for me (1)
- Somewhat true for me (2)
- Somewhat false for me (3)
- Very false for me (4)

Criticism or scolding hurts me quite a bit

- Very true for me (1)
- Somewhat true for me (2)
- Somewhat false for me (3)
- Very false for me (4)

When I want something I usually go all-out to get it

- Very true for me (1)
- Somewhat true for me (2)
- Somewhat false for me (3)
- Very false for me (4)

I will often do things for no other reason than that they might be fun

- ○ Very true for me (1)
- ○ Somewhat true for me (2)
- ○ Somewhat false for me (3)
- ○ Very false for me (4)

It's hard for me to find the time to do things such as get a haircut

- ○ Very true for me (1)
- ○ Somewhat true for me (2)
- ○ Somewhat false for me (3)
- ○ Very false for me (4)

If I see a chance to get something I want I move on it right away

- ○ Very true for me (1)
- ○ Somewhat true for me (2)
- ○ Somewhat false for me (3)
- ○ Very false for me (4)

I feel pretty worried or upset when I think or know somebody is angry at me

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

When I see an opportunity for something I like I get excited right away

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

I often act on the spur of the moment

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

If I think something unpleasant is going to happen I usually get pretty "worked up"

- ○ Very true for me (1)
- ○ Somewhat true for me (2)
- ○ Somewhat false for me (3)
- ○ Very false for me (4)

I often wonder why people act the way they do

- ○ Very true for me (1)
- ○ Somewhat true for me (2)
- ○ Somewhat false for me (3)
- ○ Very false for me (4)

When good things happen to me, it affects me strongly

- ○ Very true for me (1)
- ○ Somewhat true for me (2)
- ○ Somewhat false for me (3)
- ○ Very false for me (4)

I feel worried when I think I have done poorly at something important

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

I crave excitement and new sensations

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

When I go after something I use a "no holds barred" approach

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

I have very few fears compared to my friends

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

It would excite me to win a contest

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

I worry about making mistakes

- ◯ Very true for me (1)
- ◯ Somewhat true for me (2)
- ◯ Somewhat false for me (3)
- ◯ Very false for me (4)

End of Block: BIS/BAS

Start of Block: Thank you

Thank you very much ! The experiment is now over. You may leave after notifying the organisers.

End of Block: Thank you

B.8 Topics selection

Topics to choose

Start of Block: Default Question Block

User ID

____

Task ID: FYQ

Please choose **two topics** from below which you **know the least about** and **find interesting**.

▢ **Topic ID: 1** Your friend is going to China to live there. You hear about China's new social credit system and you want to tell her about it so she can be aware of it. Using the search engine provided, gather as much information so you can provide a unbiased overview of the subject. (1)

▢ **Topic ID: 2** Your spouse has to make a presentation and because of lack of time, they ask for your help. You have to make slides introducing 'Cognitive biases'. Gather as much information so you can introduce and gives examples of the subject. (2)

▢ **Topic ID: 3** Your colleague tells you about the recent WhatsApp vulnerability. Using the search engine provided, find out about the origin of the hack and data collected so you can convince your friends of the risks with using instant messaging apps. (3)

▢ **Topic ID: 4** You are watching the news and you hear about the danger of painkillers and their role in the US Opioid crisis. One of your friends is currently studying in the United States and you heard him mentioning the fact he is taking pain relievers. You start having some concerns for him. Using the search engine provided, find out about what is the 'US opioid crisis' so you can justify your concerns to your friend and explain the consequences of this crisis. (4)

▢ **Topic ID: 5** You recently purchased a Huawei smartphone and your friend mentioned about the recent issues the company had with the USA. You start having some doubts about your purchase. Using the search engine provided, find out about the state of the relations between the USA and the company Huawei and what is the issue. (5)

☐ **Topic ID: 6** In December 2018, Sudan has seen a series of protests to overthrow Omar Al-Bashir who was the president at the time. In April 2019, Omar Al-Bashir was overthrown. However, the protests continued. Using the search engine provided, find out about the current state of Sudan (why the protests continue and who are the different parties involved), gather as much information so you can provide a unbiased overview of the subject. (6)

☐ **Topic ID: 7** New alternatives to meat, such as Beyond Meat and Impossible Foods, have shown up to market. Recent reviews have shown promising potential but also concerns when it comes to their ingredients, some consider these alternatives as unhealthy. Using the search engine provided, gather as much information so you can provide a unbiased overview of the subject. (7)

End of Block: Default Question Block